



Trabajo fin de grado:

Gráficos de control univariantes según el promedio y la desviación típica del rango

Autor: Inés Peirats de Castro

Tutor: Ignacio Cascos Fernández

Departamento: Estadística

Grado: Ingeniería en Tecnologías
Industriales



Título: Gráficos de control univariantes según el promedio y la desviación típica del rango.

Autor: Inés Peirats de Castro

Tutor: Ignacio Fernández Cascos

Tribunal:

Presidente: Velilla Cerdán, Santiago

Secretario: Fernández Rebollo, Fernando

Vocal: Monge Alcázar, Miguel Ángel

Acto de defensa realizado el 7 de Julio de 2016 en la Escuela Politécnica Superior de la Universidad Carlos III de Madrid en Leganés.



TFG: Gráficos de control univariantes según el promedio y la desviación típica del rango

Resumen:

En este trabajo se implementan dos gráficos de control de procesos. Estos permiten monitorizar las variaciones de escala y localización de los datos de un proceso de fabricación en un solo gráfico.

El primer gráfico fue propuesto por Chao y Cheng [1] y se basa en un estadístico que combina linealmente el cuadrado de la media y la varianza de una muestra. Está basado en una suposición de normalidad.

El segundo gráfico Shewhart-Cucconi [2] se basa en los rangos de ordenación de los datos para crear un gráfico para el estadístico de Cucconi que cuantifica las alteraciones en media y dispersión. Sirve para cualquier tipo de distribución de datos.

Los resultados obtenidos se comparan entre ellos y con las gráficas clásicas de Shewhart \bar{X} y S.

Se implementan simulaciones usando los dos programas obtenidos para evaluar su rendimiento obteniendo los ARL analizando datos con diferentes distribuciones de probabilidad.

Palabras clave: gráfico 2-D, gráfico SC, estadístico de Cucconi, Shewhart, ARL

Abstract

In this project two statistical control charts are implemented. We can control the variations of scale and location using a single chart thanks to them.

The first graph, proposed by Chao and Cheng [1], uses the linear combination of the square of the mean and the variation. It's based on the assumption of normal distributed data.

The second graph Shewhart-Cucconi [2] uses the ranks of the data studied to create a chart for the Cucconi statistic that quantifies the variations in scale and location. All types of probability distributions can be used.

The results obtained are compared between them and with the Shewhart charts \bar{X} y S.

Different simulations are run using both programs to evaluate the performance. The ARL are found and analyzed using data with different probability distributions.

Key-words: 2-D chart, SC chart, Cucconi statistic, Shewhart, ARL

Índice:

	Página
Resumen	4
Abstract	5
Índice de Figuras	8
Índice de Tablas	9
<u>1-Introducción:</u>	10
1-1. Introducción y Objetivos.....	11
1-2. Entorno estadístico R.....	11
1-3. Etapas de desarrollo del proyecto.....	12
1-4. Estructura del documento.....	12
<u>2-Control de procesos:</u>	14
2-1. SPC: Control Estadístico de Procesos.....	15
2-2. Herramientas Básicas del Control de Calidad.....	16
2-3. Distribución Normal.....	18
2-3.1. QQ-plot.....	18
2-3.2. Histograma.....	20
2-3.3. Test de Shapiro-Wilk.....	21
<u>3-Graficos de control:</u>	22
3-1. Introducción a los Gráficos de Control.....	23
3-2. ARL: Average Run Length.....	24
3-3. Gráfico para la media \bar{X}	24
3-4. Gráfico para la dispersión: R y S.....	26
3-5. Gráfico CUSUM.....	29
3-6. Gráfico EWMA.....	30
3-7. Gráfico Multivariante	31
<u>4- Gráfico 2-D:</u>	33
4-1. Gráficos 2-D.....	34
4-2. Implementación gráfico	34
4-3. Implementación en R.....	36
4-4. Ejemplo.....	37
4-5. Comparación y análisis de resultados.....	39
<u>5-Gráfico Shewhart-Cucconi</u>	43
5-1. Shewhart-Cucconi	44
5-2. Implementación gráfico SC.....	44
5-3. Implementación en R.....	46
5-4. Ejemplo.....	46
5-5. Comparación y análisis de los resultados.....	48



<u>6-Análisis de Rendimiento</u>	50
6-1. Implementación en R.....	51
6-1.1. Distribución Normal.....	52
6-1.2. Distribución de Laplace.....	53
<u>7-Conclusiones</u>	55
<u>8-Referencias</u>	58
<u>9-Anexos</u>	59

Índice Figuras

	Página
<i>Figura 1:</i> Histograma datos aleatorios.....	16
<i>Figura 2:</i> Ejemplo de diagrama de Pareto.....	17
<i>Figura 3:</i> Función de densidad $N(0,1)$	18
<i>Figura 4:</i> Gráfico QQ-plot datos normales.....	19
<i>Figura 5:</i> Gráfico QQ-plot datos distribución aleatoria.....	20
<i>Figura 6:</i> Histograma distribución normal.....	21
<i>Figura 7:</i> Test de Shapiro-Wilk en R.....	21
<i>Figura 8:</i> Gráfica X-bar.....	26
<i>Figura 9:</i> Gráfica R.....	27
<i>Figura 10:</i> Gráfica S.....	29
<i>Figura 11:</i> Gráfica CUSUM.....	30
<i>Figura 12:</i> Gráfica EWMA.....	31
<i>Figura 13:</i> QQ-plot para datos históricos diámetro anillos de pistón.....	37
<i>Figura 14:</i> Gráfico 2-D para datos históricos.....	38
<i>Figura 15:</i> Gráfico 2-D para nuevos datos.....	39
<i>Figura 16:</i> Resumen datos fuera de control gráfico 2-D.....	39
<i>Figura 17:</i> Gráfico Xbar para $\alpha=0.0027$	40
<i>Figura 18:</i> Gráfico S para $\alpha=0.0027$	41
<i>Figura 19:</i> Gráfico SC datos históricos.....	47
<i>Figura 20:</i> Gráfico SC datos nuevos.....	48
<i>Figura 21:</i> Resumen datos fuera de control para SC.....	48
<i>Figura 22:</i> Comparación gráfico SC y 2-D.....	49
<i>Figura 23:</i> Ejemplo simulación para 2-D con $N(0,1)$	51
<i>Figura 24:</i> Distribución $L(0,1)$	53



Índice Tablas

	Página
<i>Tabla 1:</i> Paquetes de datos de R usados.....	12
<i>Tabla 2:</i> ARL para distribución $N(\mu, s)$	52
<i>Tabla 3:</i> ARL para distribución $L(\mu, s)$	54
<i>Tabla 4:</i> Constantes para calcular los límites de control.....	61



Capítulo 1: Introducción

En este capítulo se van a presentar los objetivos del proyecto. Además se muestran las diferentes etapas que se han llevado a cabo para la realización del proyecto así como la estructura del documento.

1-Introducción:

1-1. Introducción y objetivos

El presente trabajo consiste en la realización del software necesario para la realización de dos programas para la representación de los gráficos de control propuestos por Chao y Cheng [1] y por Chowdury, Mukherjee y Chakraborti [2]. Esta implementación se lleva a cabo usando el entorno estadístico libre R que permite la construcción de gráficos de control usando las funciones y herramientas de su base de datos.

Los gráficos de control son una herramienta clave en los procesos de fabricación para controlar la calidad del producto realizado. Su uso permite detectar anomalías de forma temprana en los procesos productivos.

Los gráficos clásicos propuestos por Shewhart para efectuar un control de proceso analizan una sola variable y hacen falta dos gráficos para poder controlar de forma eficaz la media y escala del proceso. Existen numerosas alternativas a estos gráficos.

El objetivo principal de este proyecto es crear dos gráficos de control univariantes que combinen localización y escala en un solo gráfico.

El gráfico 2-D calcula un estadístico estandarizado usando la media y la desviación típica de los datos, mientras que el gráfico SC se basa en los rangos de ordenación de los datos para crear un gráfico con una variación del estadístico de Cucconi.

1-2. Entorno estadístico R

Para la realización de los gráficos de este proyecto se usa el software libre de estadística R. Este entorno estadístico permite programar y graficar usando una base de funciones propias muy amplia que se encuentran en diferentes paquetes descargables para su utilización. También permite compartir funciones creadas en paquetes subiéndolos a una base de datos compartida. Para ejecutar los programas hay que usar comandos que se introducen en la consola.

Para poder realizar este trabajo de fin de grado tuve que aprender a programar en este entorno estadístico desde cero ayudándome de los tutoriales básicos que ofrece la misma web así como de los recursos que ofrece el Departamento de Estadística de la universidad. Las similitudes con la programación en C hicieron algo más fácil esta tarea puesto que anteriormente había programado en el software Dev C++ para una asignatura de la carrera.

Para implementar los diferentes programas y funciones para elaborar los gráficos de control presentados en los artículos se ha hecho uso de varios paquetes ya existentes en el repositorio oficial de R:

Tabla 1: Paquetes de datos de R usados

Paquete “ qcc ” [3] (quality control charts)	Este paquete permite crear gráficos de control de Shewhart para datos continuos y variables y gráficos Cusum y EWMA para datos continuos. Además realiza análisis de control de calidad. Contiene los datos pistonrings usados para los ejemplos de este trabajo.
Paquete “ nleqslv ” [4]	Resuelve sistemas de ecuaciones no lineares usando los métodos de Broyden o Newton.
Paquete “ smoothmest ” [5]	Permite crear distribuciones de Laplace o doble-exponencial.

1-3. Etapas de desarrollo del proyecto

Para la realización del proyecto en primer lugar se analizaron los gráficos de control clásicos ya existentes.

En segundo lugar se aprendió a usar el entorno estadístico R para así poder programar las funciones necesarias para la realización de los gráficos puesto que no se tenía ninguna formación previa en este lenguaje de programación.

En tercer lugar se estudiaron los artículos objeto de este TFG y se implementaron las funciones necesarias para su realización.

Por último una vez que ya se obtuvieron los programas para los dos gráficos se procedió a la realización de múltiples simulaciones para analizar y comparar los resultados.

1-4. Estructura del documento:

A continuación se presenta un breve resumen de lo que se puede encontrar en cada capítulo:

Capítulo 2: En este capítulo se presenta brevemente el origen del control de procesos (SPC). También se explica en que consiste haciendo especial énfasis a la variabilidad de in proceso.

- Capítulo 3: En este apartado se habla de lo que es un gráfico de control. Una vez explicada esta herramienta se explican algunos de los gráficos de control clásicos (\bar{X} , S, R, CUSUM, EWMA y multivariante).
- Capítulo 4: En este capítulo se estudia el gráfico 2-D de Chao y Cheng [1]. Se explica cómo se lleva a cabo la implementación del programa en R y se comparan los resultados con los gráficos de Shewhart.
- Capítulo 5: En este capítulo se implementa y estudia el gráfico Shewhart-Cucconi. Se presenta un ejemplo y se comparan los resultados con los encontrados con el gráfico 2-D.
- Capítulo 6: En este capítulo se realizan dos test de rendimiento para comprobar la eficacia de los gráficos implementados en R. Se realizan simulaciones usando diferentes distribuciones para obtener los ARL.
- Capítulo 7: Este apartado presenta las conclusiones de este trabajo y una visión personal de los objetivos conseguidos.
- Capítulo 8: Referencias y bibliografías usadas
- Capítulo 9: Anexos necesarios para realizar este proyecto. Se incluyen los programas implementados en R para la creación de los gráficos de control y el cálculo de los ARL.



Capítulo 2: Control de procesos

En este capítulo se presenta brevemente el origen del control de procesos (SPC). También se explica en que consiste haciendo especial énfasis a la variabilidad de in proceso.

2-Control de procesos

2-1. SPC: Control Estadístico de Procesos

Para poder llegar a entender lo que es un control de proceso hay que saber lo que la calidad representa. La calidad es una medida que permite determinar si un producto, proceso o cosa esta libre de defectos. Gracias a ella se pueden comparar dos propiedades iguales y determinar su variabilidad, es decir cuanta diferencia hay entre uno y otro. La calidad pretende mantener una uniformidad y unos estándares dentro del proceso para satisfacer en todo momento las necesidades del cliente. Un proceso va a estar siempre afectado por factores que varían como la mano de obra, las máquinas, el entorno, los materiales.

En los años veinte Bell Telephone, empresa de telecomunicaciones, buscaba soluciones a los inconvenientes que causaban las reparaciones de los equipos subterráneos que instalaban. Pusieron en marcha un sistema de inspección para evitar mandar productos con defectos a los clientes. Este sistema, aunque efectivo, suponía grandes costes a la compañía y daba lugar a una baja eficiencia de la producción. Walter Shewhart [6], físico, ingeniero, estadístico y trabajador de la compañía estudió los fallos que se iban encontrando en los equipos y determinó que existían dos clases diferentes: fallos debido a variaciones comunes y fallos debidos a variaciones especiales:

-Las causas comunes de variabilidad se definen como propias a la naturaleza del proceso. Estas variaciones son causadas por las 6Ms: material, maquinaria, mano de obra, métodos, medición y medio ambiente. Son difíciles de eliminar debido a su aleatoriedad y no afectan en la calidad final del proceso ya que representan pequeños cambios. Shewhart se dio cuenta que el intentar eliminar estas causas producía más variabilidad y fallos que los que ya se tenían.

-Las causas especiales son variaciones no previstas debidas a un cambio del proceso fuera de lo normal. Algunos ejemplos de causas que pueden provocar estos fallos son: la utilización de materiales distintos a los especificados, el fallo de una máquina o la mala manipulación por parte de un operario. Esta variabilidad tiene como resultado un proceso que no cumple con la calidad y los estándares marcados. Estas causas pueden ser detectadas y eliminadas con mayor facilidad.

En los controles de proceso se busca que las variaciones provengan sólo de causas comunes.

De las observaciones y el trabajo llevado a cabo por Shewhart se puso en marcha lo que hoy se conoce como SPC: Control Estadístico de Procesos. El SPC son una serie de herramientas que permiten controlar la calidad allí donde se crea. Usa datos generales en vez de comparar la variabilidad dato a dato. El SPC es muy útil puesto que se puede utilizar en cualquier proceso donde el producto final tenga características mensurables. Al estudiar un proceso en más detalle se aumenta el conocimiento sobre este y su mejora resulta más fácil.

2-2. Herramientas Básicas del Control de Calidad

Para reducir al máximo las anomalías y la variabilidad presentes en los controles de proceso se usan 7 herramientas estadísticas básicas:

- Histograma: representación gráfica en forma de barras de la variable analizada. Muestra la frecuencia de los resultados ordenados por categorías. Se usa para visualizar de manera rápida la dispersión de los datos.

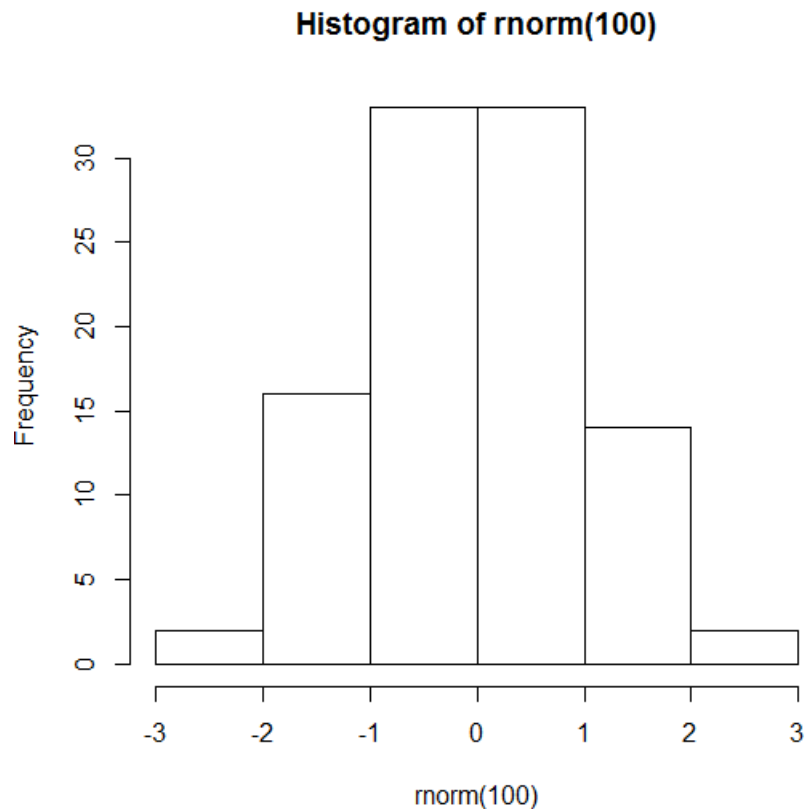


Figura 1: Histograma datos aleatorios

- Diagrama de Pareto: histograma de la variable analizada donde las frecuencias están ordenadas de mayor a menor. Permite asignar prioridades a los problemas más importantes (con más frecuencia) de aquellos que son triviales.

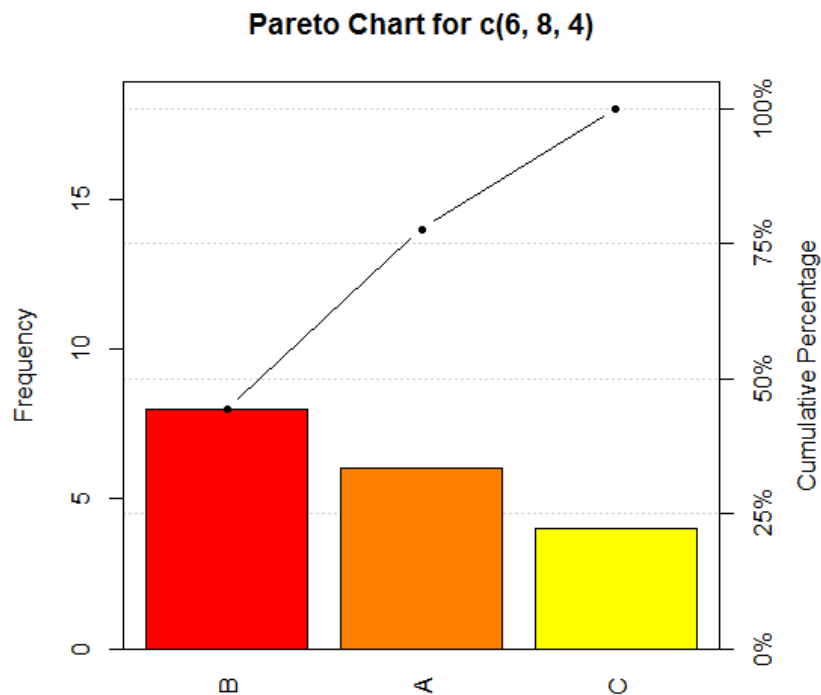


Figura 2: Ejemplo de diagrama de Pareto

- Diagrama de dispersión: estudian la relación entre dos variables. Si existe correlación las dos variables serán directa o inversamente proporcionales.
- Diagrama Causa-Efecto: se presenta mediante una flecha horizontal el problema a analizar. Haciendo uso del brain-storming se añaden flechas con las posibles causas de la variabilidad de la característica estudiada.
- Hoja de verificación: se recolectan las frecuencias de la característica estudiada y se presentan en categorías en forma de tablas. Sirve para registrar resultados y tendencias.
- Diagrama de flujo: se clasifican los datos en subconjuntos, estratos, teniendo en cuenta factores comunes. Haciendo uso de los histogramas o diagramas de Pareto se grafican los datos. Se comparan los grupos entre ellos para detectar las causas que más afectan la variabilidad.
- Gráfico de control: representan valores sucesivos de la característica estudiada. En el Capítulo 3 se estudian en detalle.

2-3. Distribución Normal

Los datos usados en los gráficos clásicos siguen una distribución Normal(μ, σ). La función de densidad de unos datos que siguen una distribución normal tiene forma de campana y es simétrica con respecto a su media.

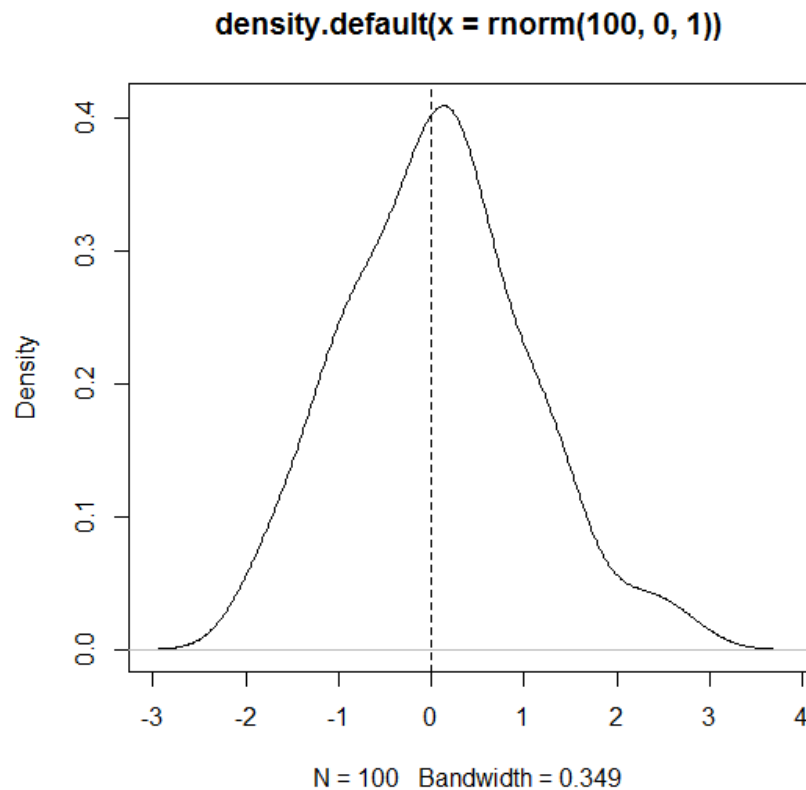


Figura 3: Función de densidad $N(0,1)$

En la figura 3 se puede observar la función de densidad de media 0 y desviación típica 1. La función de densidad tiene un máximo en la media, crece hasta la media y decrece después. El área dentro de la curva es igual a la unidad y es igual a 0.5 a cada lado de la media.

El uso de datos normales permite la obtención de resultados exactos sin necesidad de usar métodos de aproximación.

Dada una serie de datos existen diversas maneras de determinar si se puede asimilar a una distribución normal.

2-3.1 QQ-plot:

El uso de gráficos probabilísticos de normalidad, como el qq-plot, permite comparar la distribución de una serie de datos con la distribución Normal.

Con este gráfico se quiere estudiar la hipótesis nula H_0 : los datos de la serie provienen de una distribución Normal.

Para construir este gráfico, para una serie de m datos, se ordenan los datos de forma ascendente. A cada dato ordenado se le asigna el cuantil teórico de una distribución Normal con la probabilidad p_i igual a:

$$p_i = \frac{i - \frac{3}{8}}{m + \frac{1}{4}}$$

Una vez se tienen todos los datos se representa los puntos de ordenadas (cuantil teórico, cuantil observación) en un gráfico. Se añade una línea de tendencia para la mejor visualización.

Si los datos están alineados no se rechaza la hipótesis H_0 de normalidad.

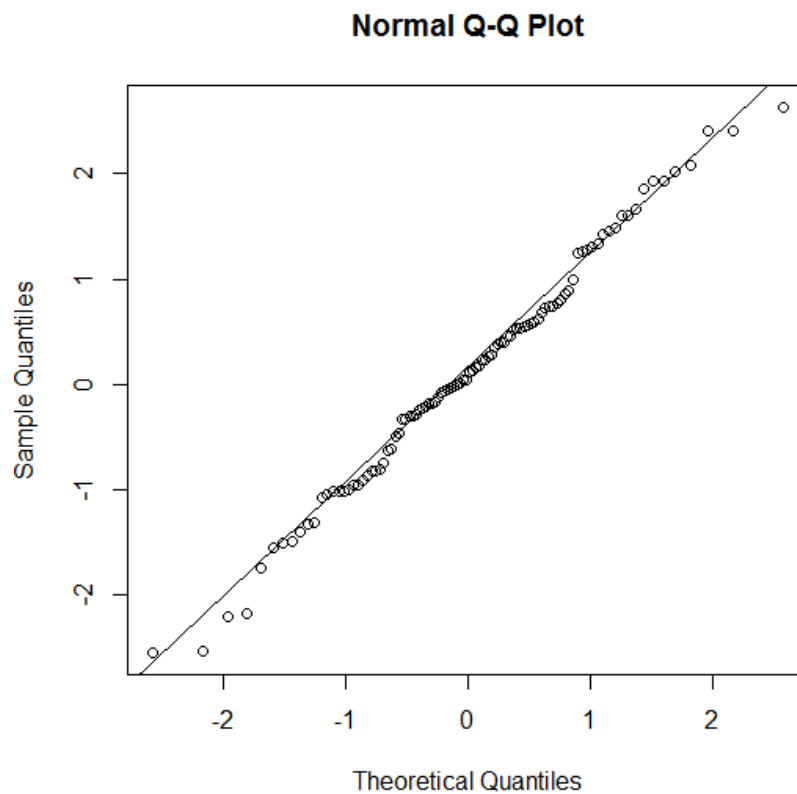


Figura 4: Gráfico QQ-plot datos normales

Comando: `qqnorm(x<-rnorm(100,0,1) ; qqline(x)`

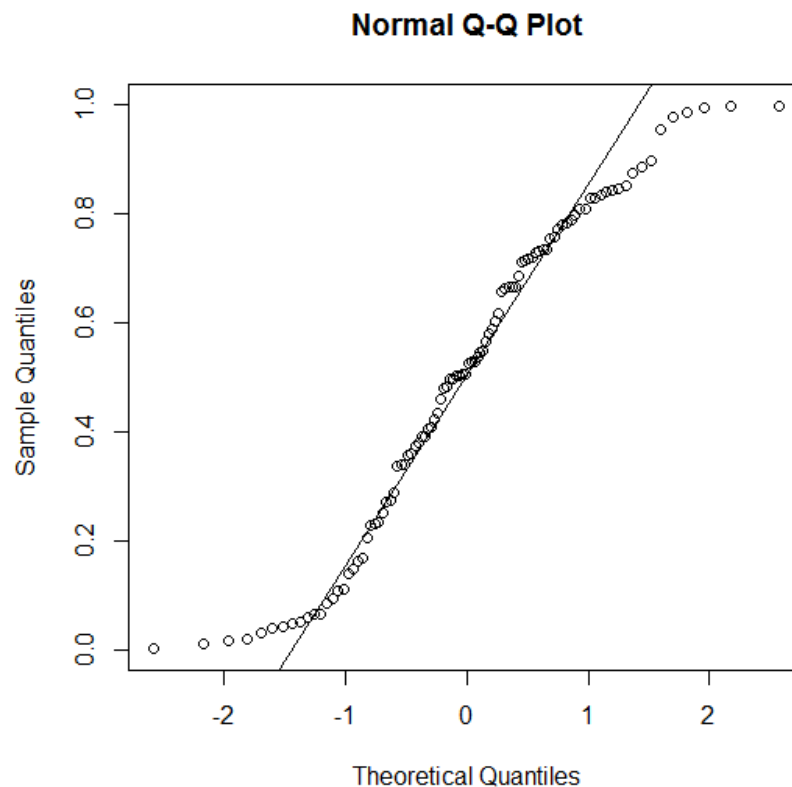


Figura 5: Gráfico *QQ*-plot datos distribución aleatoria

Comando: `qqnorm(x<-runif(100)) ; qqline(x)`

En la figura 4 los datos están alineados, los datos de la serie siguen una distribución normal. En la figura 5 los datos no están alineados y se alejan de la recta por lo que se rechaza la hipótesis nula.

2-3.2 Histograma

Otra forma gráfica de determinar la normalidad de una serie de datos se basa en el uso de su histograma. Si se tiene una serie lo suficientemente grande (más de 20 datos) los histogramas son una buena representación gráfica de la dispersión de los datos.

La hipótesis nula de normalidad se acepta si el histograma de los datos tienen forma de campana de Gauss y es simétrico en torno a μ .

Usando una serie de 100 datos normales podemos observar en R la forma de su histograma.

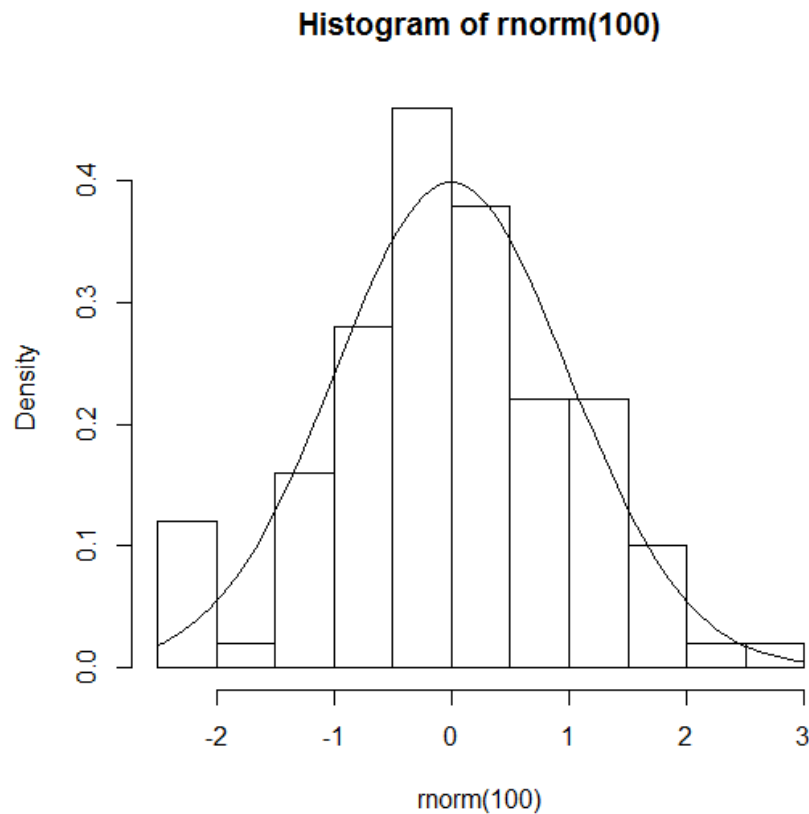


Figura 6: Histograma distribución normal

Comando: `hist(rnorm(100),freq=F); curve(dnorm(x),add=TRUE)`

2-3.3 Test de Shapiro-Wilk:

El test de Shapiro-Wilk es un test de contraste de normalidad usado para muestras de tamaño m inferior a 50. Si se quieren datos normales entonces se busca aceptar la hipótesis nula de normalidad H_0 .

Para realizar este test hay que calcular la media μ y la varianza muestral S . Se ordenan los datos en orden ascendente y se calcula un estadístico.

Si el p-valor obtenido es mayor a 0.05 no se puede rechazar la hipótesis nula.

Ejemplo con datos en R para una distribución uniforme de 100 datos:

```
Shapiro-Wilk normality test

data:  runif(100)
W = 0.95838, p-value = 0.003086
```

Figura 7: Test de Shapiro-Wilk en R

Comando: `shapiro.test(runif(100))`

El p-valor es inferior a 0.05 por lo que rechazamos la hipótesis de normalidad.



Capítulo 3: Gráficos de control

En este apartado se habla de lo que es un gráfico de control. Una vez explicada esta herramienta se explican algunos de los gráficos de control clásicos (\bar{X} , S, R, CUSUM, EWMA y multivariante).

3-Gráficos de Control:

3-1 Introducción a los gráficos de control

Los gráficos de control son la herramienta usada por el SPC para controlar la variabilidad de un proceso a lo largo del tiempo y su calidad. Detectan rápidamente la aparición de causas especiales y son usados para mejorar los procesos.

Estos gráficos representan diferentes estadísticos de muestras tomadas siguiendo un orden cronológico. Los datos usados para graficar el proceso pueden ser tomados de forma individual o en subgrupos de tamaño muestral n .

Un gráfico de control está compuesto por una línea central (LC) que corresponde al promedio del estadístico y dos líneas horizontales, una a cada lado del LC, que indican los límites de control superior (LCS) e inferior (LCI).

La primera fase de la realización de un gráfico de control es el análisis de los datos históricos. Los límites se determinan teniendo en cuenta el nivel de confianza σ fijado. En los gráficos propuestos por Shewhart [7], con datos que siguen una distribución normal, se toma un nivel de significación $\alpha = 0.0027$ correspondiente a 3σ : los límites estarán situados a una distancia tres veces la desviación estándar de la media establecida teniendo un 99.73% de probabilidades de que las muestras estén entre los límites de control. Esto quiere decir que hay una probabilidad del 0.27% que un punto se encuentre fuera de los límites y se detecte una falsa alarma. La variación del nivel sigma varía la probabilidad de falsa alarma en los gráficos clásicos. A menor nivel sigma mayor la probabilidad de que un punto se encuentre dentro de los límites.

Una vez se tienen las líneas de límites se inicia la segunda fase del control monitorizando las muestras futuras. Los puntos se juntan mediante líneas para que los cambios se vean con más facilidad. En este punto se da lugar al análisis de los datos recolectados.

El control estadístico de proceso comprueba los datos obtenidos. El principal objetivo de los gráficos es detectar los puntos que caen fuera de los límites de control. Cuando esto pasa el control manda una señal y hay que investigar la alarma en busca de una variación de tipo especial.

Los gráficos de control también permiten detectar rachas de puntos que pueden causar problemas en la calidad del proceso. Algunas de estas rachas son:

- Dos de tres puntos seguidos del mismo lado de la línea central y a una distancia mayor de 2σ de esta.
- Cuatro de cinco puntos seguidos del mismo lado de la LC y a una distancia mayor de 1σ de esta.
- Si más de 8 puntos están del mismo lado de la LC.
- Si los puntos muestran una tendencia, ciclos o una no aleatoriedad aun estando dentro de los límites.

Para supervisar un control de procesos usando las técnicas propuestas por Shewhart de manera eficaz hay que estudiar dos gráficos a la vez: uno para la localización y otro para la media.

3-2. ARL: Average Run Length

La rapidez de detección de un gráfico se puede medir gracias al ARL_o (average run length) que representa el número de muestras en promedio que el control de proceso establecido tiene que analizar antes de la primera muestra fuera de control. El subíndice indica que el proceso está en control.

Sabiendo α el ARL_o se calcula como:

$$ARL_o = \frac{1}{\alpha}$$

De esta manera si se usa un nivel 3-sigma se tiene $\alpha = 0.0027$ y el ARL_o será igual a 370. Esto significa que, en promedio, la observación 370 va a estar fuera de los límites.

Para controlar el rendimiento de un proceso se mide el ARL y se compara con el teórico.

A continuación se presentan varios gráficos clásicos de control SPC [7]. Se tomara siempre un nivel 3σ para calcular los límites de control.

Se toman los datos “*pistonrings*” propuestos en el paquete de datos “*qcc*” para todos los ejemplos de este capítulo. Los datos usados se encuentran en el Anexo 1[3].

3-3. Gráfico para la media \bar{X}

El gráfico de control \bar{X} presenta la media aritmética de las muestras tomadas durante un proceso industrial agrupadas en subgrupos. Las muestras tienen un mismo tamaño n y se asume que siguen una distribución normal según el teorema del límite central (la media aritmética de una muestra lo suficientemente grande de valores aleatorios e independientes sigue una distribución normal). Esta gráfica muestra la variación de las medias muestrales alrededor de un valor para la media establecida.

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

En el caso que la media μ y la desviación típica σ sean conocidas los límites de control se calculan directamente. Si se toma un nivel 3 sigma, suponiendo una distribución normal, el 99,7% de las medias muestrales estarán contenidas entre los límites de control siguientes:

$$LCS_{\bar{X}} = \mu + 3 \frac{\sigma}{\sqrt{n}} = \mu + A\sigma$$

$$LC_{\bar{X}} = \mu$$

$$LCI_{\bar{X}} = \mu - 3 \frac{\sigma}{\sqrt{n}} = \mu - A\sigma$$

El coeficiente A tiene el 3 incorporado.

En los procesos industriales es común no conocer μ y σ . En esos casos hay que usar unos datos históricos para determinar el promedio de las medias $\bar{\bar{x}}$ y la media de las desviaciones típicas \bar{s} y usar las formulas siguientes:

$$LCS_{\bar{X}} = \bar{\bar{x}} + \frac{3\bar{s}}{c_4\sqrt{n}} = \bar{\bar{x}} + A_1\bar{s}$$

$$LC_{\bar{X}} = \bar{\bar{x}}$$

$$LCI_{\bar{X}} = \bar{\bar{x}} - \frac{3\bar{s}}{c_4\sqrt{n}} = \bar{\bar{x}} - A_1\bar{s}$$

c_4 es una constante que depende del tamaño de la muestra n .

La desviación estándar de la muestra s mide el grado de dispersión de los datos muestrales para cada subgrupo y se calcula como la raíz cuadrada de la varianza de los datos.

$$s = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}}$$

Ejemplo gráfico \bar{X} :

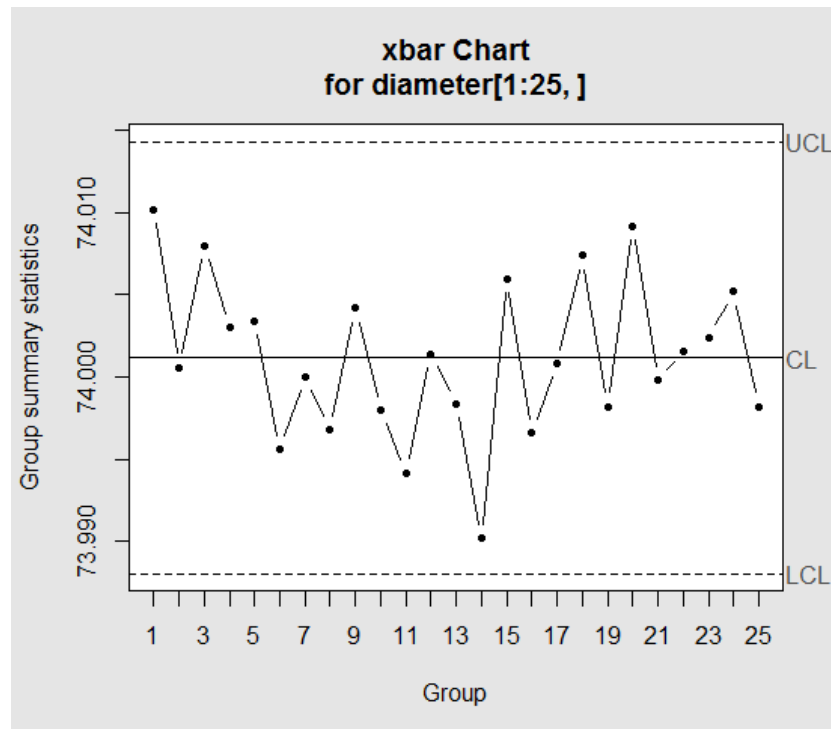


Figura 8: Gráfica X-bar

Comando: `qcc(diameter[1:25,],type="xbar",newdata=diameter[26:40,])`

3-4. Gráfico para la dispersión: R y S

A la hora de realizar un control de procesos es necesario complementar el gráfico de control de los promedios con un gráfico de dispersión. Estos gráficos representan la variabilidad del proceso.

Los gráficos de rango R presentan la diferencia entre la mayor y la menor observación de cada muestra i :

$$R = \max(R_i) - \min(R_i)$$

Este gráfico sólo utiliza dos valores de cada muestra lo que puede causar falsas alarmas si se usan muestras de gran tamaño ($n > 8$) puesto que se obvian los demás datos.

Los límites de control siendo μ y σ conocidos se calculan usando las ecuaciones:

$$LCS_R = (d_2 + 3d_3)\sigma = D_2\sigma$$

$$LC_R = d_2\sigma$$

$$LCI_R = (d_2 - 3d_3)\sigma = D_1\sigma$$

Los límites de los gráficos R usando datos históricos se encuentran con las siguientes fórmulas:

$$LCS_R = \left(1 + 3 \frac{d_3}{d_2}\right) \bar{R} = D_4 \bar{R}$$

$$LC_R = \bar{R}$$

$$LCI_R = \left(1 - 3 \frac{d_3}{d_2}\right) \bar{R} = D_3 \bar{R}$$

Donde \bar{R} es el promedio de los rangos de los datos históricos.

Ejemplo gráfico R:

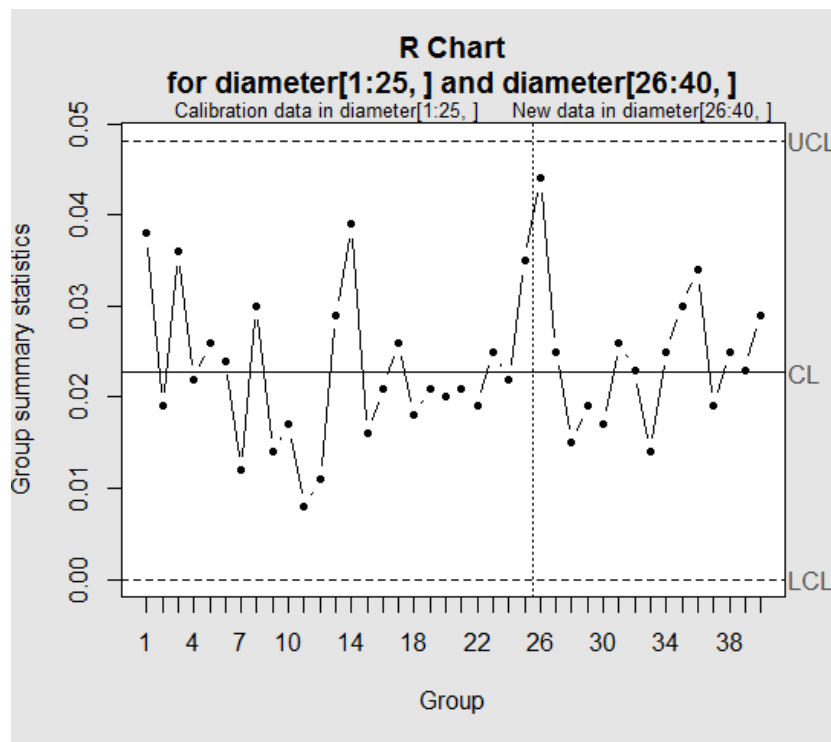


Figura 9: Gráfica R

Comando: `qcc(diameter[1:25,],type="R",newdata=diameter[26:40,])`

El gráfico de desviaciones estándar S utiliza todos los valores de la muestra por lo que es más preciso aunque menos utilizado debido a la mayor complejidad de los cálculos usados para su trazado.

Para esta gráfica se calcula la desviación estándar insesgada de cada subgrupo de tamaño n usando la fórmula siguiente:

$$S = \sqrt{\frac{\sum_{i=1}^n X_i - \bar{X}}{n - 1}}$$

Para los límites de control con datos de media y varianza conocidos se utilizan las formulas:

$$LCS_S = \left(c_4 + 3\sqrt{1 - c_4^2} \right) \sigma = B_6 \sigma$$

$$LC_{S^*} = c_4 \sigma$$

$$LCI_S = \left(c_4 - 3\sqrt{1 - c_4^2} \right) \sigma = B_5 \sigma$$

Los límites usando datos históricos se encuentran gracias a las fórmulas:

$$LCS_S = \left(1 + \frac{3}{c_4} \sqrt{1 - c_4^2} \right) \bar{S} = B_4 \bar{S}$$

$$LC_S = \bar{S}$$

$$LCI_S = \left(1 - \frac{3}{c_4} \sqrt{1 - c_4^2} \right) \bar{S} = B_3 \bar{S}$$

\bar{S} corresponde a la media de las desviaciones de los datos históricos.

Ejemplo gráfico S

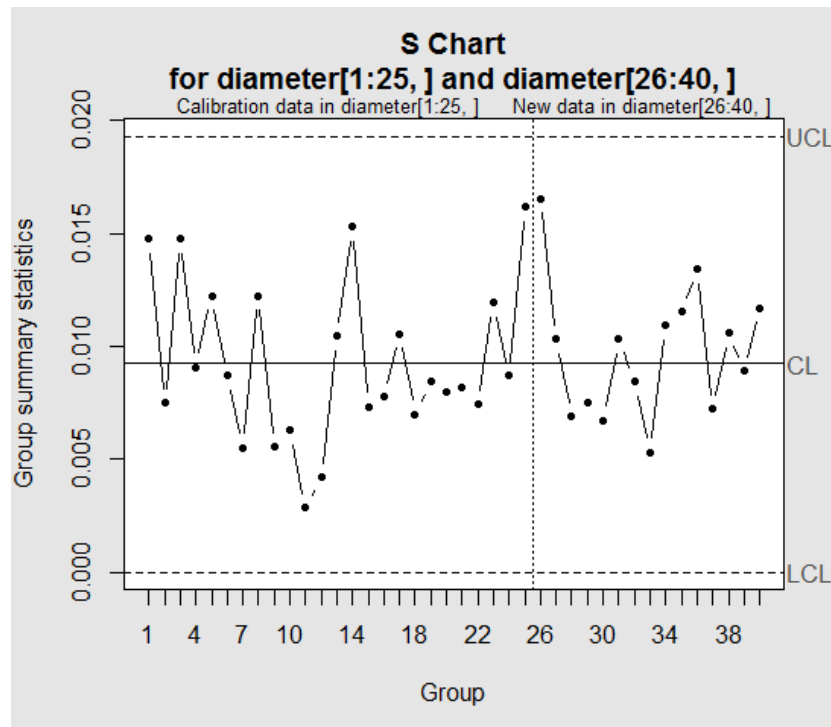


Figura 10: Gráfica S

Comando: `qcc(diameter[1:25,],type="S",newdata=diameter[26:40,])`

3-5. Gráfico CUSUM

El gráfico CUSUM (cumulative sum control chart) para el control estadístico de procesos muestra la suma acumulada de las desviaciones con respecto a la media.

Si se tienen m subgrupos de n datos cada uno y una media μ la suma acumulada se calcula mediante la fórmula siguiente:

$$S_m = \sum_{i=1}^m (\bar{X}_i - \mu)$$

Este gráfico detecta cambios pequeños más rápidamente que los gráficos tradicionales de Shewhart.

El gráfico advierte de un problema si se detecta una tendencia creciente o decreciente. Mientras los datos se mantengan centrados en la media μ los valores acumulados se mantendrán cerca de 0.

Para construir el gráfico CUSUM es necesario fijar un valor d para la desviación mínima que se quiere detectar. Los límites de control quedarían por lo tanto como:

$$LCS_{CUSUM} = \mu + d$$

$$LC_{CUSUM} = \mu$$

$$LCI_{CUSUM} = \mu - d$$

Ejemplo gráfico CUSUM:

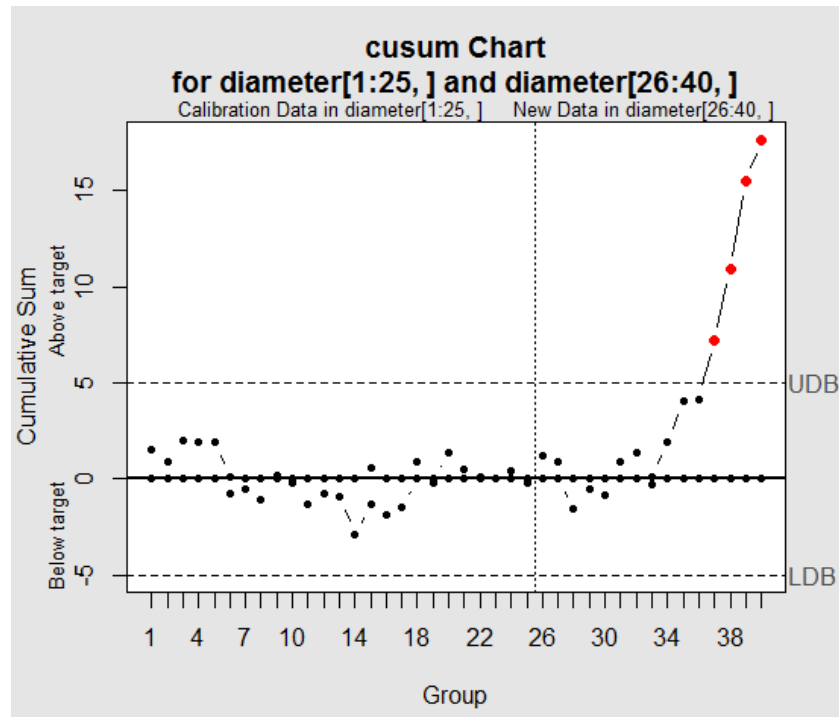


Figura 11: Gráfica CUSUM

Comando: `cusum(diameter[1:25,],newdata=diameter[26:40,])`

3-6. Gráfica EWMA

Los gráficos EWMA son diagramas de medias móviles con pesos exponenciales. Los puntos muestran información de sus antecedentes lo que permite detectar las desviaciones rápidamente.

Si llamamos α al peso, teniendo m subgrupos de n muestras cada uno, el estadístico EWMA viene dado por la fórmula:

$$Z_m = \alpha \bar{X}_m + (1 - \alpha)Z_{i-1}$$

$$Z_0 = \frac{1}{m} \sum_{i=1}^m \bar{X}_i$$

Los valores de α pequeños se usan para detectar cambios pequeños más rápidos.

Los límites de control se establecen mediante las ecuaciones:

$$LCS_{EWMA} = Z_0 + \frac{3\sigma}{\sqrt{n}} \sqrt{\frac{\alpha}{1-\alpha}} [1 - (1-\alpha)^{2m}]$$

$$LC_{EWMA} = Z_0$$

$$LCI_{EWMA} = Z_0 - \frac{3\sigma}{\sqrt{n}} \sqrt{\frac{\alpha}{1-\alpha}} [1 - (1-\alpha)^{2m}]$$

Ejemplo gráfico EWMA:

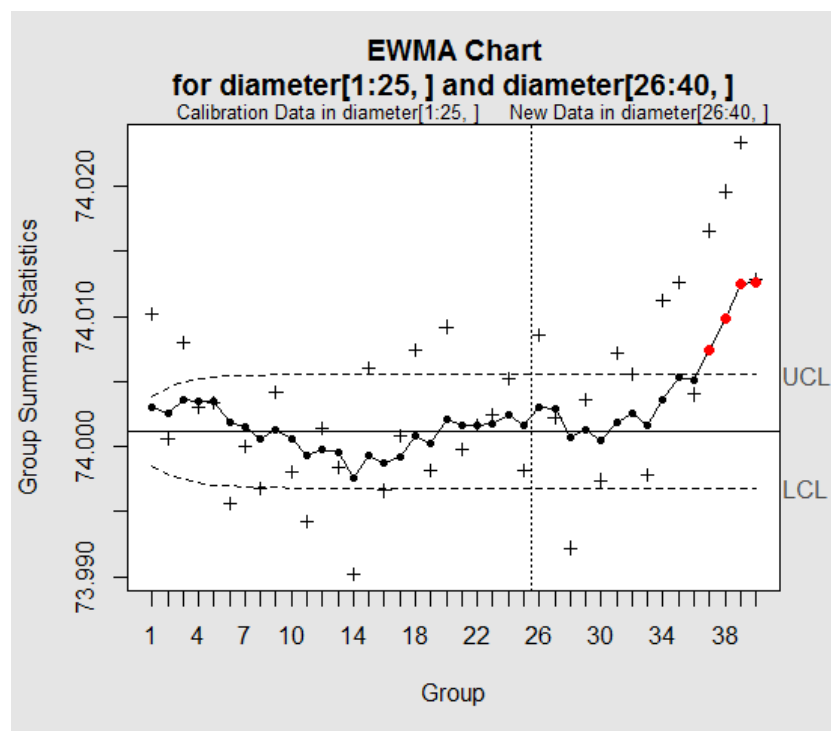


Figura 12: Gráfica EWMA

Comando: `ewma(diameter[1:25,],newdata=diameter[26:40,])`

3-7. Gráfico Multivariante:

Los gráficos multivariantes presentan más de una variable para cada observación.



Para representar los datos se usan matrices con n filas, donde n es el número de datos en cada muestra, y p columnas, donde p es el número de variables estudiadas.

Las constantes $A, A_1, D_2, D_1, D_3, D_4, B_4, B_6, B_5, B_4$ y B_3 dependen del tamaño muestral n y del nivel de significación α y viene recogidas en la tabla del Anexo 2 [8].



Capítulo 4: Gráfico 2-D

En este capítulo se estudia el gráfico 2-D de Chao y Cheng [1]. Se explica cómo se lleva a cabo la implementación del programa en R y se comparan los resultados con los gráficos de Shewhart.

4-Gráfico 2-D:

4-1. Gráficos 2-D:

En los controles de procesos que siguen los gráficos propuestos por Shewhart se analizan muestras de datos con distribución normal $N(\mu, \sigma)$ usando dos gráficos diferentes para la localización y la dispersión. Estos gráficos siguen una secuencia cronológica y los estadísticos se representan en función del tiempo en el que han sido obtenidos. En su trabajo, “On 2-D control charts”, los estadísticos Min-Te Chao y Smiley W. Cheng proponen una forma diferente de monitorizar un proceso de producción con datos univariantes usando un sólo gráfico. De esta forma la localización y la dispersión estarían representados por un mismo estadístico estandarizado en un gráfico bi-dimensional.

Hay dos formas de aproximar un gráfico 2-D usando el promedio \bar{X} y la desviación S :

Aproximación 1: Usar datos con coordenadas (\bar{X}, S) , los datos no siguen un orden cronológico pero no se pierden grados de libertad. Los límites de control pasarían a ser representados por un área cerrada.

Aproximación 2: Representar \bar{X} y S con un mismo estadístico T en función del tiempo.

Para la primera aproximación presentan el caso de un gráfico con un área de control circular. Los puntos bajo control se encontrarían dentro del área delimitada por una elipse, o círculo si se usa una gráfica a escala.

Para la segunda aproximación, la que se va a implementar en el entorno estadístico R, hay que calcular un estadístico que incluya la dispersión y la localización.

El estadístico T usado por los autores para la creación del gráfico 2-D no es único. Para cualquier constante $b > 0$ se puede encontrar T_b para representar diagramas 2-D exactos y similares:

$$T_b = n(\bar{X} - \mu)^2 + bS_i^2$$

Donde n es el número de muestras por observación y μ se encuentra usando los datos históricos.

4-2. Implementación gráfico:

Una vez se acepta la hipótesis de normalidad de los datos históricos se puede llevar a cabo los cálculos para la construcción del gráfico 2-D óptimo propuesto en el artículo.

Se usan k muestras de tamaño n como datos históricos y se calcula su promedio total:

$$\bar{\bar{X}} = \frac{1}{kn} \sum_{i=1}^{kn} X_i$$

A partir de la media se calcula la varianza y la desviación típica S_p de los datos:

$$S_p^2 = \frac{1}{kn-1} \sum_{i=1}^{kn} (X_i - \bar{\bar{X}})^2$$

$$S_p = \sqrt{S_p^2}$$

Para cada muestra, $j=1,2,\dots,kn$, se define x^* como la diferencia del valor de la muestra y la media total dividida entre la desviación típica.

$$x_j^* = \frac{x_j - \bar{\bar{X}}}{S_p}$$

Para cada grupo i , con $i=1,2,\dots,k$ se calcula ε como la media de las x^* y t como la desviación estándar:

$$\gamma_i = \frac{1}{n} \sum_{z=1}^n x_{iz}^*$$

$$t_i = \frac{1}{n} \sum_{z=1}^n (x_{iz}^* - \gamma_i)^2 \quad \text{con } z=1,2,\dots,n$$

Una vez se tienen todos los datos se calcula el estadístico η^2 propuesto por Min-Te Chao y Smiley W.Cheng que permite centrar los datos tomados en el origen y adimensionarlos, los datos carecen de unidad de medida asociada. Este paso permite la estandarización de los datos para poder compararlos entre ellos.

$$\eta_i^2 = \left(1 - \frac{1}{n}\right) t_i^2 - 2 \left(1 - \frac{2}{n}\right) \log(t_i) \quad \text{con } i=1,2,\dots,k.$$

El estadístico T se calcula para cada grupo como:

$$T_i = \gamma_i^2 + \eta_i^2$$

Se representa en un gráfico los puntos con coordenadas (i, T_i) , con $i=1,2,\dots,k$.

En este tipo de gráfico de control de proceso sólo hay que limitar los datos por arriba, el límite de control inferior será igual a cero. Para encontrar el LC se pueden usar tablas propuestas en el artículo o efectuando los cálculos que se proponen en su anexo. El LC depende del número de muestras n en cada grupo y del nivel de significación α .

Para que el control estadístico sea fiable los datos históricos tienen que estar bajo control.

Una vez se tienen los estadísticos de los datos históricos y el límite de control LC se pueden efectuar los cálculos para los nuevos datos usando la media total $\bar{\bar{X}}$ y la desviación estándar S_p encontrados con los datos históricos. De esta forma se calcula el estadístico T para los nuevos datos.

Se grafican las muestras que se van obteniendo y se comparan con el límite de control.

El gráfico manda una señal de alarma si $T_i < LC$. En ese caso el control de proceso está fuera de control hay que investigar los problemas que han podido causar la variación por causas especiales de la muestra i y solucionarlos en caso de que no sean una falsa alarma.

4-3. Implementación en R:

Los cálculos presentados en el apartado anterior se implementan en R para tener un programa que realice un control de proceso con un único gráfico 2-D.

La función `musigma(datas,n,k,alpha,newdata)` creada, Anexo 4, tiene por datos de entrada:

datas	datos históricos con distribución normal
n	número de muestras en cada grupo
k	número de grupos de datos históricos
alpha	nivel de significación fijado
newdata	Muestras de datos que se quieren controlar de tamaño n

La función `CL(n,alpha)`, Anexo 3, calcula el límite de control para un tamaño muestral n y un nivel de significación α fijados.

La función tiene por salida un único gráfico de control que presenta el estadístico T de los datos históricos y los nuevos datos tomados en función del número de muestra.

Además el programa indica las muestras que están fuera de control tanto gráficamente, puntos en rojo, como por escrito.

4-4. Ejemplo

Para probar el funcionamiento del programa se usan los datos *pistonrings* incluidos en el paquete “qcc”. La variable que se va a controlar es el diámetro de unos anillos de pistón.

En total se tienen 40 muestras de las cuales las 25 son datos históricos, $k=25$, cada muestra tiene $n=5$ observaciones.

Lo primero que tenemos que hacer es estudiar la serie de datos históricos para poder aceptar o rechazar la hipótesis de normalidad. Para ello vamos a usar el qq-plot.

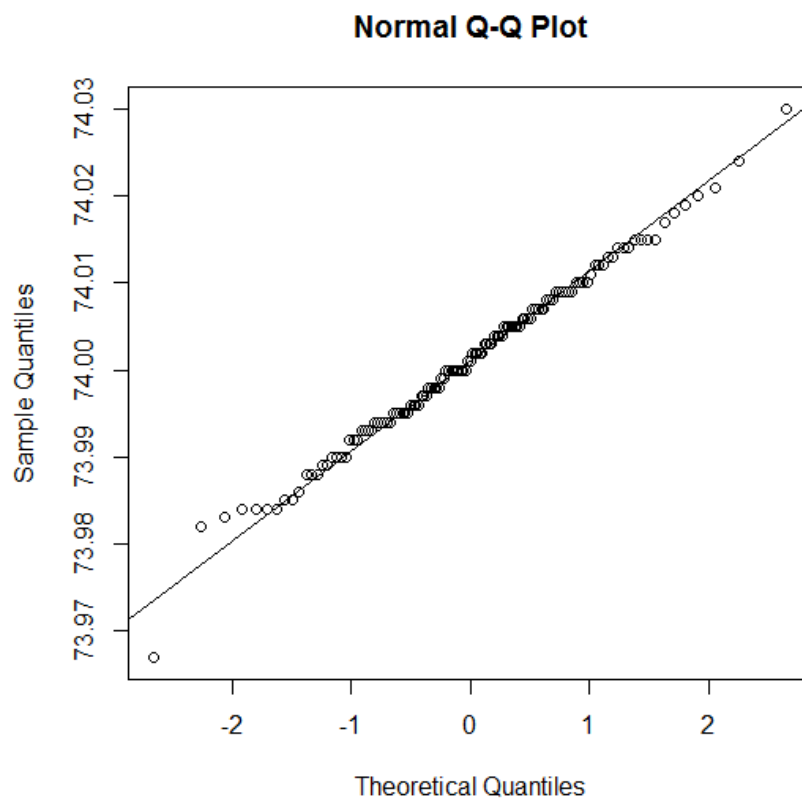


Figura 13: QQ-plot para datos históricos diámetro anillos de pistón

Comando: `qqnorm(diameter[1:25,]) ; qqline(diameter[1:25,])`

Los puntos están alineados en una recta por lo que no se puede rechazar la hipótesis de normalidad.

Tenemos evidencias estadísticas de que los datos históricos tomados para el diámetro de los anillos de pistón pueden asimilarse a una distribución normal.

El siguiente paso es analizar los datos históricos y encontrar el límite de control.

Usando la función $CL(n, \alpha)$ se encuentra el límite de control $LC=3.0851$ para un nivel de significación $\alpha = 0.0027$.

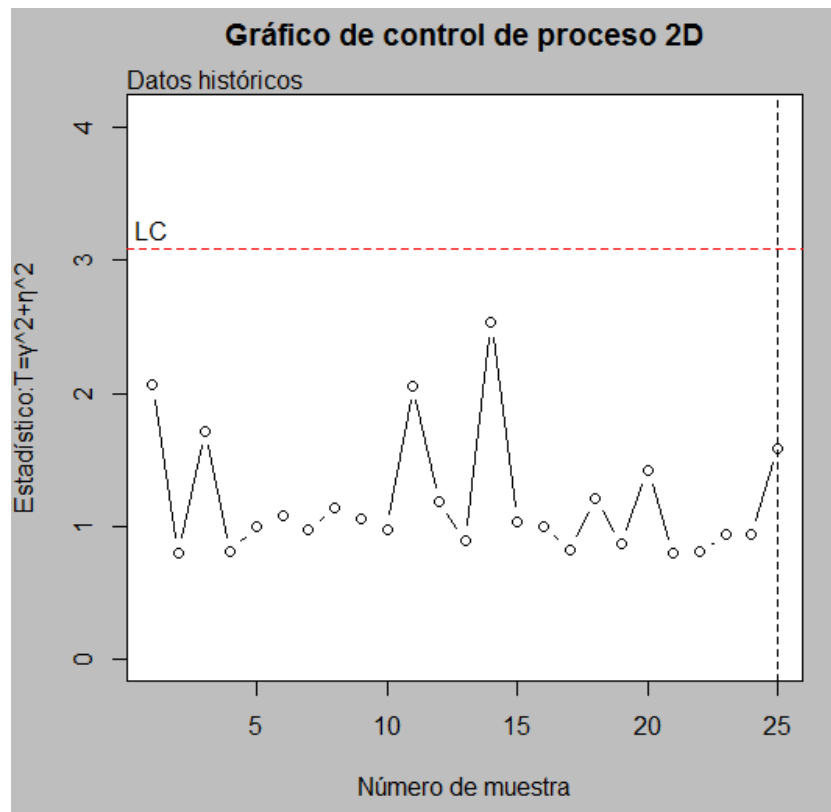


Figura 14: Gráfico 2-D para datos históricos

Comando: `musigma(diameter[1:25,],5,25,0.0027)`

Todos los puntos se sitúan por debajo del límite de control por lo que no hay alarmas de que las muestras estén fuera de control.

La gráfica se puede usar para estudiar las muestras nuevas tomadas en el proceso. Se añaden 15 observaciones con 5 muestras cada una.

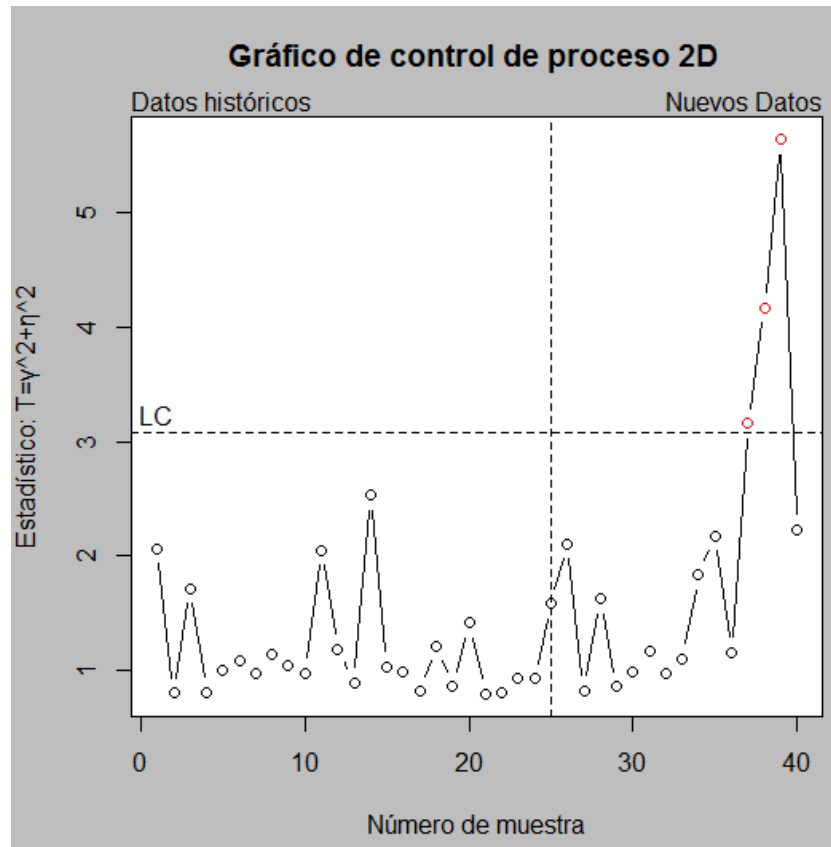


Figura 15: Gráfico 2-D para nuevos datos

```
[1] "La muestra 37 está fuera de control"
[1] "La muestra 38 está fuera de control"
[1] "La muestra 39 está fuera de control"
```

Figura 16: Resumen datos fuera de control gráfico 2-D

Comando: `musigma(diameter[1:25,],5,25,0.0027,diameter[26:40,])`

Se puede observar gráficamente que hay 3 observaciones por encima del límite de control que corresponden a las muestras 37, 38 y 39. El proceso por lo tanto no está bajo control. Las observaciones presentan pequeñas variaciones mientras se mantienen bajo control y una gran variación una vez el proceso está fuera de control.

Si se toma un $\alpha = 0.0054$, el límite de control pasa a ser $LC=2.8101$. Las muestras fuera de control siguen siendo las mismas.

4-5. Comparación y análisis de los resultados

Usando los mismos datos se obtienen las gráficas \bar{X} propuestas por Sthewhart. Las figuras 17 y 18 muestran los gráficos de control \bar{X} y S respectivamente para un nivel $\alpha = 0.0027$.

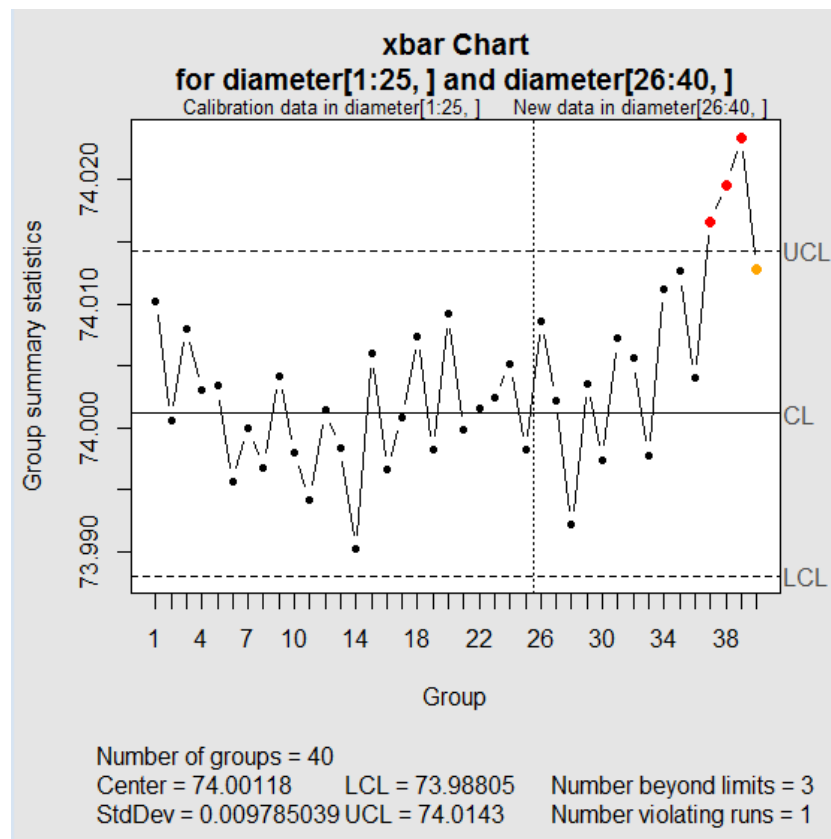


Figura 17: Gráfico Xbar para $\alpha=0.0027$

Comando: `qcc(diameter[1:25,],type="xbar",newdata=diameter[26:40,])`

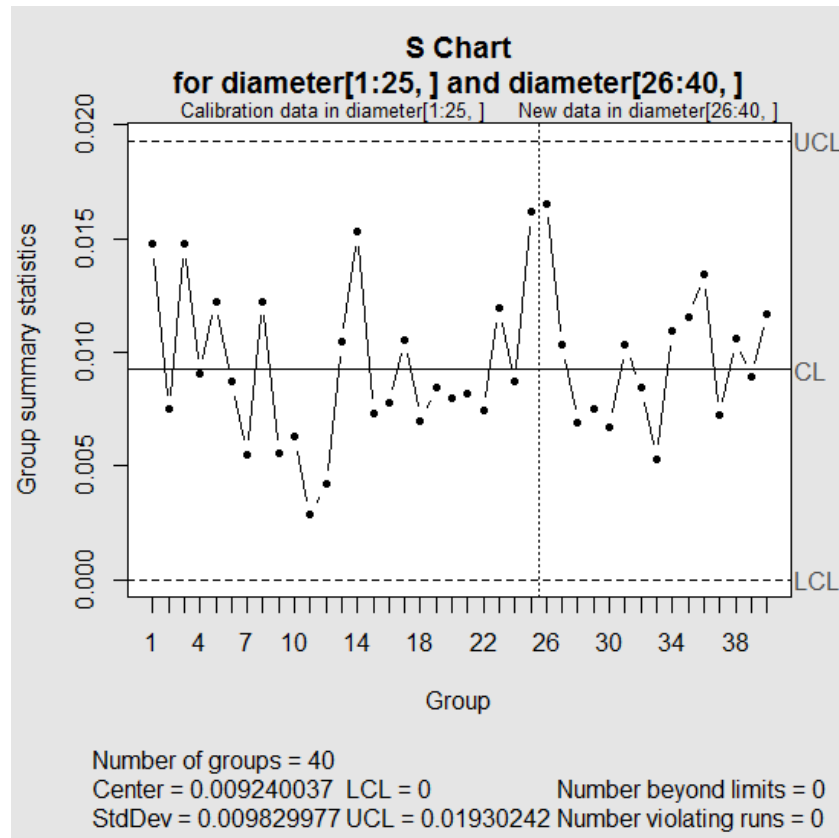


Figura 18: Gráfico S para $\alpha=0.0027$

Comando: `qcc(diameter[1:25,],type="S",newdata=diameter[26:40,])`

Se puede observar que la gráfica de dispersión S no presenta ningún punto fuera de los límites de control. La gráfica \bar{X} presenta 3 puntos fuera de control que corresponden a las mismas observaciones que el gráfico 2-D (37, 38 y 39). Se observa también un número elevado de puntos del mismo lado de la LC a partir de la observación 34 lo que manda una señal de alarma.

Usando el artículo de Min-Te Chao y Smiley W. Cheng se ha realizado la programación de un gráfico de control. Este gráfico considera a la vez las variaciones de la localización y de la dispersión.

Comparando los resultados con las gráficas de Shewhart se puede observar que la gráfica detecta las variaciones aleatorias de forma más conservadora con cambios menos marcados. La variación entre los puntos bajo control y los puntos fuera de los límites es más visible y clara en el gráfico 2-D lo cual es una ventaja y facilita la rápida detección de los fallos en el proceso.

En el caso de los gráficos de Shewhart es necesario analizar los dos gráficos \bar{X} y S a la vez. El análisis de uno sólo puede llevar a error. En el ejemplo presentado el gráfico S no presenta ningún punto fuera de los límites, si se estudiase sólo el proceso con ese gráfico podría ser considerado bajo control. Sin embargo el gráfico X si muestra puntos fuera de control y una tendencia creciente clara.



El gráfico de control 2-D necesita de una calculadora o programa de ordenador para su realización y es más complejo de implementar que otros gráficos.



Capítulo 5: Gráfico Shewhart-Cucconi

En este capítulo se implementa y estudia el gráfico Shewhart-Cucconi. Se presenta un ejemplo y se comparan los resultados con los encontrados con el gráfico 2-D.

5-Gráfico Shewhart-Cucconi

5-1. Shewhart-Cucconi

La mayoría de los controles de proceso más comunes necesitan para ser validos la aceptación de la hipótesis de normalidad de los datos históricos. El gráfico de control SC basado en el estadístico de Cucconi que se va a implementar permite el uso de datos con una distribución continua y libre.

Los datos no-paramétricos o de distribución libre no suponen ninguna distribución de probabilidad. Para poder realizar pruebas no-paramétricas se necesita utilizar datos independientes y aleatorios.

El control de procesos de Cucconi usa los rangos y contra-rangos de ordenación al cuadrado para resolver el problema de juntar localización y escala en un mismo gráfico.

5-2. Implementación gráfico SC

El artículo propone una serie de pasos para la construcción del gráfico.

El primer paso es recolectar m datos históricos U_m del proceso de control que se quiera analizar. Se ordenan los datos en orden creciente y se les asigna su rango de ordenación.

Las nuevas observaciones con distribución continua V_n de tamaño n se recogen. Se crea una matriz combinada ordenada con los datos históricos y los n nuevos datos. Se definen los estadísticos T_1 y S_1 . Para cada nueva observación T_1 es la suma de rangos de sus n muestras en la matriz combinada y S_1 la suma de los rangos al cuadrado.

$$T_1 = \sum_{k=1}^N k I_k$$

$$S_1 = \sum_{k=1}^N k^2 I_k$$

N es la posición del dato en la matriz combinada ($N=n+m$). I_k es igual a 0 si el dato pertenece a los U_m datos y 1 si es una observación V_n .

Una vez calculados estos dos estadísticos se define S_2 para cada nueva observación como:

$$S_2 = \sum_{k=1}^N (N+1-k)^2 I_k = n(N+1)^2 - 2(N+1)T_1 + S_1$$

Se estandarizan los datos U_m y V_n para poder compararlos con las siguientes ecuaciones:

$$U = \frac{6S_1 - n(N+1)(2N+1)}{\sqrt{\frac{mn}{s}(N+1)(2N+1)(8N+11)}}$$

$$V = \frac{6S_2 - n(N+1)(2N+1)}{\sqrt{\frac{mn}{s}(N+1)(2N+1)(8N+11)}}$$

El coeficiente de correlación ρ entre U y V sólo depende del tamaño N de la matriz combinada. Si se cogen siempre el mismo número de muestras en cada observación este dato permanece constante y es igual a:

$$\rho = \frac{2(N^2 - 4)}{(2N+1)(8N+11)} - 1$$

El estadístico de Cucconi se calcula para cada observación usando los datos U y V y el coeficiente de correlación ρ :

$$C = \frac{U^2 + V^2 - 2\rho UV}{2(1 - \rho^2)}$$

Para que el control estadístico sea fiable los datos históricos deben estar bajo control. Para comprobar esto se calculan los C_m para las muestras de datos históricos.

Una vez calculados hay que buscar el límite de control LC de valor H que depende del tamaño de los datos históricos m , de n y del nivel de significación se establezca para el proceso.

Se grafican los puntos de coordenadas (j, C_j) con $j=1,2,\dots,m$ y el límite de control horizontal de valor H .

Si todos los C_j son inferiores a LC los datos históricos están bajo control y se puede empezar a analizar nuevos datos. En caso contrario habría que investigar donde está el problema y solucionarlo antes de empezar el proceso de control.

Los estadísticos C para los nuevos datos se calculan y grafican. Cuando uno de los puntos excede el límite de control el proceso manda una señal de alarma y el proceso se declara

fuera de control. En ese caso hay que efectuar tests para la observación para determinar si es un problema de localización, de escala o una falsa alarma.

5-3. Implementación en R

Los cálculos presentados en el apartado anterior para tener un programa que realice un gráfico SC usando datos no paramétricos para el control de un proceso.

La función `rango(sample, m, n, alpha, test)` creada tiene por datos de entrada

sample	Datos históricos con distribución no-paramétrica
m	Número de datos históricos totales
n	Número de muestras en cada observación
alpha	Nivel de significación fijado
test	Nuevos datos

La función `CLrango(m, n, alpha)` permite obtener mediante unas tablas el límite de control para un tamaño muestral n , un número m de datos históricos y un nivel de significación α fijados.

La función `rango` tiene como salida un único gráfico que representa el estadístico C en función del número de observación y el límite de control.

Los datos fuera de control están representados en rojo en el gráfico. El programa también avisa del número de la observación fuera de control por escrito.

5-4. Ejemplo

Para probar el programa se usan los datos *pistonrings* incluidos en el paquete “*qcc*”.

Se tienen $m=125$ datos históricos y 15 observaciones con $n=5$ muestras como datos nuevos.

El primer paso es analizar los datos históricos y encontrar el límite de control.

En este caso se ha usado un nivel de significación $\alpha=0.0027$ y el límite de control es igual a $LC=5.77$.

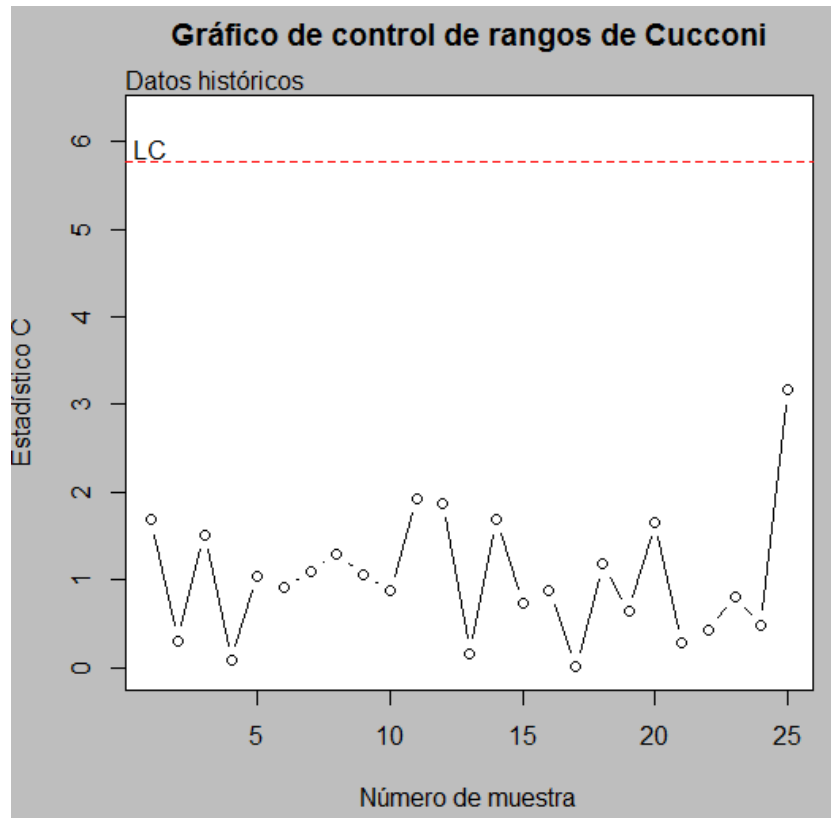


Figura 19: Gráfico SC datos históricos

Comando: rango(diameter[1:25,],125,5,0.0027)

Se puede observar que todos los puntos están por debajo del límite de control LC por lo que el proceso está bajo control y se puede usar para observaciones futuras.

Se añaden las 15 nuevas observaciones al proceso.

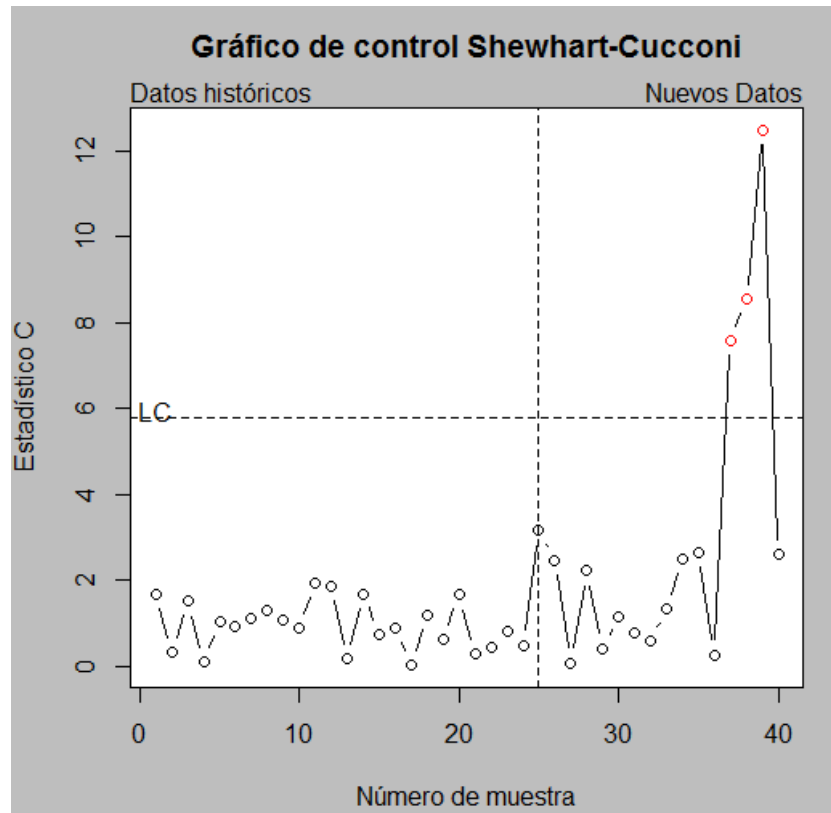


Figura 20: Gráfico SC datos nuevos

```
[1] "La muestra 37 está fuera de control"
[1] "La muestra 38 está fuera de control"
[1] "La muestra 39 está fuera de control"
```

Figura 21: Resumen datos fuera de control para SC

Comando: `rango(diameter[1:25,],125,5,0.0027,diameter[26:40,])`

Se puede observar que hay tres puntos por encima del LC. Las observaciones 37, 38 y 39 mandan una señal de alarma, están fuera de control. El cambio entre la observación 36 y la 37 es brusco, la alarma está marcada y es fácil de detectar. Mientras están bajo control las observaciones presentan pequeñas variaciones.

5-5. Comparación y análisis de los resultados

Basándonos en el artículo de S.Chowdhury, A.Mukherjee y S. Chakraborti se ha implementado un gráfico de control usando el estadístico de Cucconi que detecta las variaciones de escala y localización a la vez usando los rangos ordenados de las diferentes observaciones.

En la Figura 22 podemos comparar el gráfico 2-D y el SC para los mismos datos.

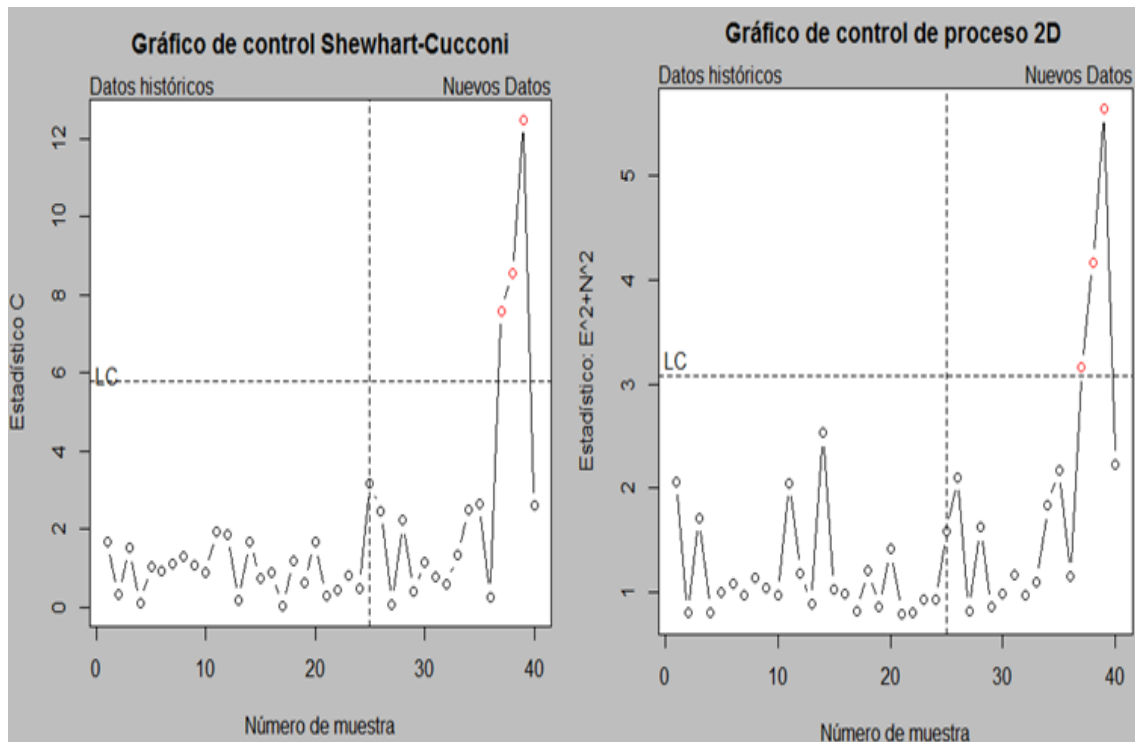


Figura 22: Comparación gráfico SC y 2-D

Las observaciones por encima del LC son las mismas (37, 38, 39). Sin embargo en el gráfico SC las anomalías aparecen más marcadas, la variación entre las observaciones bajo control y fuera de control es bastante mayor. Se puede ver que la observación 37 está muy cerca del LC en el caso del gráfico 2-D mientras que en el gráfico SC se aleja considerablemente de éste.

Los datos bajo control presentan menos fluctuaciones en el gráfico SC. Este gráfico es más conservador que el otro. Las alarmas por superación del LC son más evidentes pero las alarmas por falta de aleatoriedad o por tendencia son difíciles de apreciar.

El gráfico de Shewhart-Cuconni usa cálculos más sencillos y por lo tanto es más fácil y rápido de implementar que el 2-D.



Capítulo 6: Análisis de Rendimiento

En este capítulo se realizan dos test de rendimiento para comprobar la eficacia de los gráficos implementados en R. Se realizan simulaciones usando diferentes distribuciones para obtener los ARL.

6-Análisis de Rendimiento

6-1. Simulaciones en R

Para testar el rendimiento de los programas implementados en R se van a obtener los ARL para datos con diferentes distribuciones.

Los programas se van a ejecutar en bucle 1000 veces, mil datos históricos aleatorios de mismo tamaño, para comprobar su fiabilidad. La primera etapa del programa es comprobar que los datos históricos están bajo control. De ser así el programa crea nuevas observaciones de tamaño n hasta que una observación se encuentra por encima del LC. El número de observaciones hasta llegar a la observación fuera de control se guarda y se crean nuevos datos históricos.

Se implementan las funciones `ARLmusigma(n, k, alpha, mean1, sd1)`, Anexo 7, y `ARLrango(m, n, alpha, mean1, sd1)`, Anexo 8, usando las funciones `musigma` y `rango` como base para crear las simulaciones.

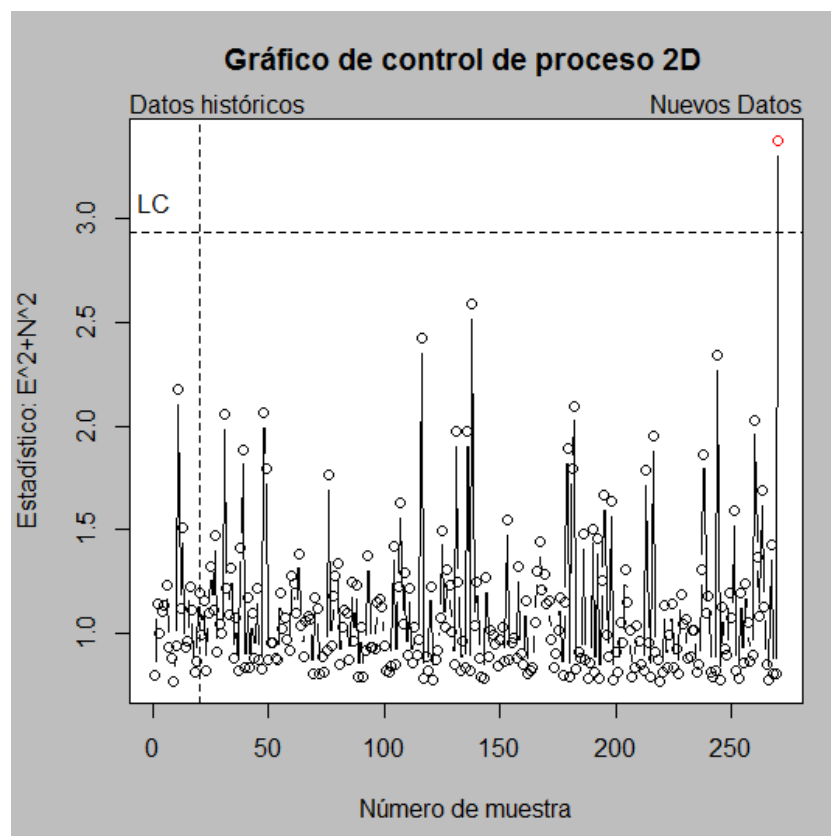


Figura 23: Ejemplo simulación para 2-D con $N(0,1)$

Comando: `ARLmusigma(5,20,0.004,0,1)`

6-1.1. Distribución Normal

La primera prueba de rendimiento se realiza usando 100 datos históricos ($k=20$ para la función 2-D y $m=100$ para la función SC). Los datos históricos siguen siempre la misma distribución normal $N(0,1)$.

Se analizan los valores encontrados para el ARL para diferentes valores de la media μ y la desviación típica s para la distribución usada para los nuevos datos. La Tabla 2 muestra los resultados obtenidos para un nivel de significación $\alpha = 0.004$ o lo que es lo mismo un ARL igual a 250.

Tabla 2: ARL para distribución $N(\mu, s)$

	2-D	SC
N(0 , 1)	254.465	258.522
N(0.25 , 1)	152.762	144.862
N(0.5 , 1)	47.019	39.01
N(1 , 1)	5.911	5.178
N(0 , 1.25)	30.963	45.128
N(0 , 0.5)	438.006	No detecta
N(0.25 , 1.25)	20.99	34.274

Se puede observar que para las simulaciones que usan nuevos datos con distribución igual a la usada para los datos históricos los ARL encontrados son muy parecidos al teórico.

Cuando se efectúa una variación en la escala, en la media μ de la distribución, los resultados encontrados para los dos gráficos son más bajos cuanto más diferencia entre la media de la distribución muestral y la de nuevos datos. La eficacia del control disminuye en más de un 90% pasando a tener una media de 5 observaciones antes de que el control mande una señal de alarma.

Cuando se establece una desviación típica de $s=0.5$ para los nuevos datos se obtiene un ARL mucho mayor que el esperado: 438 frente a un ARL teórico de 250. El proceso es menos preciso y detecta menos alarmas.

Los resultados obtenidos para los dos gráficos son muy parecidos. El gráfico 2-D parece ser más eficaz cuando se efectúa un cambio de escala, tiene valores del ARL más altos y por lo tanto la alarma es dada más tarde y el proceso está bajo control durante más

tiempo. El gráfico SC da mejores resultados cuando se efectúa un cambio de la localización o de ambos.

6-1.2. Distribución de Laplace

El segundo test de rendimiento se realiza usando distribuciones de Laplace también conocidas como doble-exponencial. Es una distribución continua definida por un parámetro de localización, su promedio μ , y un parámetro de escala s , $L(\mu, S)$.

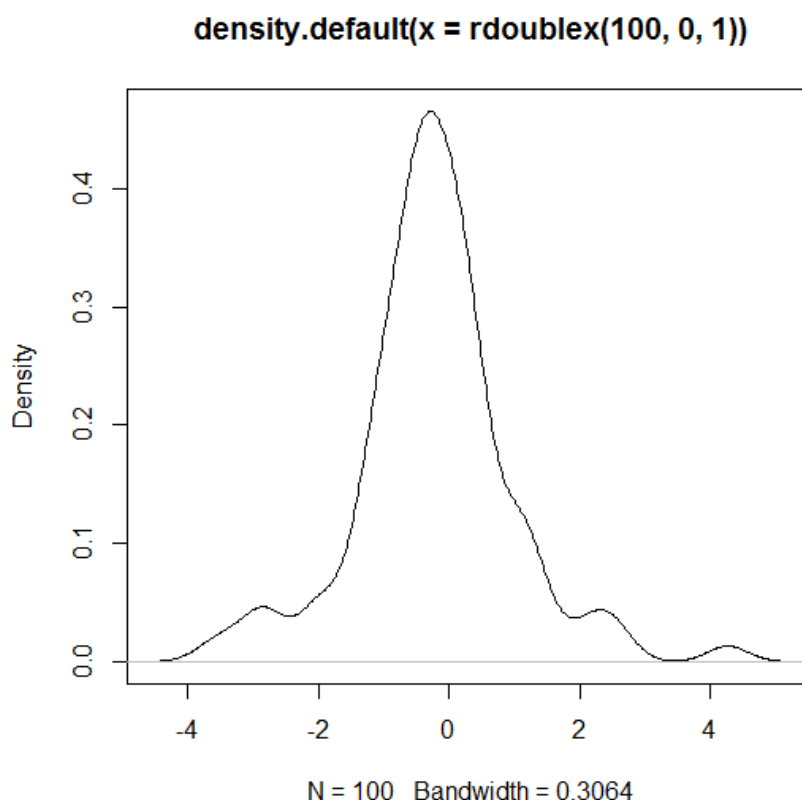


Figura 24: Distribución $L(0,1)$

Comando: `plot(density(rdouplex(100,0,1)))`

La distribución de Laplace tiene una forma más puntiaguda que la distribución Normal.

Para el segundo test de rendimiento se realizan 1000 simulaciones usando en cada una 100 datos históricos con distribución $L(0,1)$ generados aleatoriamente. La Tabla 3 recopila los ARL obtenidos usando nuevos datos con distribuciones de Laplace con valores de la media y de la desviación diferentes. El valor del nivel de significación establecido es de $\alpha = 0.004$.

Tabla 3: ARL para distribución $L(\mu, s)$

	2-D	SC
L(0 , 1)	38.434	255.796
L(0.25 , 1)	34.295	193.283
L(0.5 , 1)	27.469	94.206
L(1 , 1)	12.14	15.104
L(0 , 1.25)	12.434	67.755
L(0 , 0.5)	159.452	No detecta
L(0.25 , 1.25)	12.161	53.051

Se puede observar una gran diferencia entre los resultados obtenidos con los dos gráficos.

Para el gráfico SC los resultados son parecidos a los encontrados usando las distribuciones normales. El valor del ARL se aproxima mucho al ARL teórico de 250 cuando se usan una distribución $L(0,1)$ tanto para los datos históricos como para los nuevos datos. El valor del ARL disminuye si se mantiene el parámetro de dispersión y se aumenta la media pasando a 15 observaciones antes de estar fuera de control para una media igual a 1. Si se aumenta sólo la localización el ARL disminuye también.

En el caso del gráfico 2-D los resultados obtenidos muestran un rendimiento muy bajo. El control de procesos 2-D funciona de manera óptima para datos con distribución normal. Al usar distribuciones de Laplace el control detecta alarmas en observaciones muy tempranas y no llega en ninguno de los casos al ARL teórico.



Capítulo 7: Conclusiones

Este apartado presenta las conclusiones de este trabajo y una visión personal de los objetivos conseguidos.

7-Conclusiones

En este proyecto de fin de grado se han implementado en R dos gráficos de control de procesos. Además, se ha comparado el análisis de su rendimiento gracias a la obtención de los ARL correspondiente a cada simulación realizada.

Se ha presentado una mejora a los gráficos clásicos más usados en los procesos industriales, gráficos \bar{X} y S. El uso de un único gráfico de control para controlar la variación de escala y localización resulta en una disminución del tiempo necesario para evaluar los resultados. Esto puede suponer una reducción del tiempo del proceso y por lo tanto un aumento de su eficacia.

En el caso de los gráficos de Shewhart y 2-D los datos usados tienen que tener una distribución de probabilidad normal para que su rendimiento sea elevado y que su uso sea el adecuado. Las simulaciones realizadas muestran ARL muy bajos cuando su distribución no es normal.

Como se ha podido comprobar mediante las simulaciones el gráfico Shewhart-Cucconi es eficaz tanto con datos normales como con datos con otras distribuciones. Esto lo convierte en una herramienta muy potente y a tener en cuenta a la hora de controlar un proceso.

Aun teniendo grandes ventajas estos gráficos tienen una desventaja clara: la implementación de sus cálculos y su entendimiento no es tan inmediata como puede ser la de la construcción de un gráfico \bar{X} y S. Los operarios de planta deben ser capaces de analizar estos datos para que su utilización sea óptima.

Con este TFG se han logrado los objetivos que se habían planteado:

- Aprender a programar en el entorno estadístico R desde 0
- Analizar el uso de los gráficos de control
- Implementar y estudiar el rendimiento de dos gráficos de control

A nivel personal este proyecto me ha permitido usar los conocimientos adquiridos en los 4 años de carrera pudiendo tener una visión más práctica de ellos. También he podido comprobar como las diferentes asignaturas se entrelazan y son necesarias para poder llegar a un análisis completo del problema propuesto.



TFG: Gráficos de control univariantes según el promedio y la desviación típica del rango

Capítulo 8: Referencias

- [1] Chao, M., Cheng, S.W. *On 2-D Control Charts*. Quality Technology and Quantitative Management. 2008; Vol 5(No 3): pp. 243-261.
- [2] Chowdhury, S, Mukherjee, A, Chakraborti, S. *A New Distribution-free Control Chart for joint Monitoring of Unknown Location and Scale Parameters of Continuous Distributions*. Quality and Reliability Engineering International. 2013; doi:10.1002/qre.1488.
- [3] Scrucca, L. (2004). qcc: an R package for quality control charting and statistical process control. R News 4/1, 11-17.
- [4] Berend Hasselman (2016). nleqslv: Solve Systems of Nonlinear Equations. R package version 3.0. <https://CRAN.R-project.org/package=nleqslv>
- [5] Christian Hennig (2012). smoothest: Smoothed M-estimators for 1-dimensional location. R package version 0.1-2. <https://CRAN.R-project.org/package=smoothest>
- [6] Smith, J.L. 1. *Remembering Walter A Shewhart's Contribution to the Quality World*. Marzo 2009. Disponible en: <http://www.qualitymag.com/articles/85973-remembering-walter-a-shewhart-s-contribution-to-the-quality-world>
- [7] Gutierrez Pulido, H., De La Vara Salazar, R. *Control Estadístico de la Calidad y Seis Sigma*. (2nd ed.). México: McGraw-Hill; 2009
- [8] Hernández garza, M. *Tabla de Constantes para Gráficos de Control*. Disponible en: <https://optyestadistica.wordpress.com/2008/08/27/tabla-de-constantas-para-graficos-de-control/>



Capítulo 9: Anexos

Anexos necesarios para realizar este proyecto. Se incluyen los programas implementados en R para la creación de los gráficos de control y el cálculo de los ARL.

Anexo 1:

Datos diámetro *pistonrings* 5 muestras (columnas) para cada observación (filas)

	[,1]	[,2]	[,3]	[,4]	[,5]
1	74.030	74.002	74.019	73.992	74.008
2	73.995	73.992	74.001	74.011	74.004
3	73.988	74.024	74.021	74.005	74.002
4	74.002	73.996	73.993	74.015	74.009
5	73.992	74.007	74.015	73.989	74.014
6	74.009	73.994	73.997	73.985	73.993
7	73.995	74.006	73.994	74.000	74.005
8	73.985	74.003	73.993	74.015	73.988
9	74.008	73.995	74.009	74.005	74.004
10	73.998	74.000	73.990	74.007	73.995
11	73.994	73.998	73.994	73.995	73.990
12	74.004	74.000	74.007	74.000	73.996
13	73.983	74.002	73.998	73.997	74.012
14	74.006	73.967	73.994	74.000	73.984
15	74.012	74.014	73.998	73.999	74.007
16	74.000	73.984	74.005	73.998	73.996
17	73.994	74.012	73.986	74.005	74.007
18	74.006	74.010	74.018	74.003	74.000
19	73.984	74.002	74.003	74.005	73.997
20	74.000	74.010	74.013	74.020	74.003
21	73.988	74.001	74.009	74.005	73.996
22	74.004	73.999	73.990	74.006	74.009
23	74.010	73.989	73.990	74.009	74.014
24	74.015	74.008	73.993	74.000	74.010
25	73.982	73.984	73.995	74.017	74.013
26	74.012	74.015	74.030	73.986	74.000
27	73.995	74.010	73.990	74.015	74.001
28	73.987	73.999	73.985	74.000	73.990
29	74.008	74.010	74.003	73.991	74.006
30	74.003	74.000	74.001	73.986	73.997
31	73.994	74.003	74.015	74.020	74.004
32	74.008	74.002	74.018	73.995	74.005
33	74.001	74.004	73.990	73.996	73.998
34	74.015	74.000	74.016	74.025	74.000
35	74.030	74.005	74.000	74.016	74.012
36	74.001	73.990	73.995	74.010	74.024
37	74.015	74.020	74.024	74.005	74.019
38	74.035	74.010	74.012	74.015	74.026
39	74.017	74.013	74.036	74.025	74.026
40	74.010	74.005	74.029	74.000	74.020

Anexo 2:

Tabla 4: Constantes para calcular los límites de control

Constantes para Gráficos de Control																
n	A	A2	A3	c4	1/c4	B3	B4	B5	B6	d2	d3	1/d2	D1	D2	D3	D4
2	2.121	1.880	2.659	0.798	1.253	0.000	3.267	0.000	2.606	1.128	0.853	0.886	0.000	3.686	0.000	3.267
3	1.732	1.023	1.954	0.886	1.128	0.000	2.568	0.000	2.276	1.693	0.888	0.591	0.000	4.358	0.000	2.575
4	1.500	0.729	1.628	0.921	1.085	0.000	2.266	0.000	2.088	2.059	0.880	0.486	0.000	4.698	0.000	2.282
5	1.342	0.577	1.427	0.940	1.064	0.000	2.089	0.000	1.964	2.326	0.864	0.430	0.000	4.918	0.000	2.114
6	1.225	0.483	1.287	0.952	1.051	0.030	1.970	0.029	1.874	2.534	0.848	0.395	0.000	5.079	0.000	2.004
7	1.134	0.419	1.182	0.959	1.042	0.118	1.882	0.113	1.806	2.704	0.833	0.370	0.205	5.204	0.076	1.924
8	1.061	0.373	1.099	0.965	1.036	0.185	1.815	0.179	1.751	2.847	0.820	0.351	0.388	5.307	0.136	1.864
9	1.000	0.337	1.032	0.969	1.032	0.239	1.761	0.232	1.707	2.970	0.808	0.337	0.547	5.394	0.184	1.816
10	0.949	0.308	0.975	0.973	1.028	0.284	1.716	0.276	1.669	3.078	0.797	0.325	0.686	5.469	0.223	1.777
11	0.905	0.285	0.927	0.975	1.025	0.321	1.679	0.313	1.637	3.173	0.787	0.315	0.811	5.535	0.256	1.744
12	0.866	0.266	0.886	0.978	1.023	0.354	1.646	0.346	1.610	3.258	0.778	0.307	0.923	5.594	0.283	1.717
13	0.832	0.249	0.850	0.979	1.021	0.382	1.618	0.374	1.585	3.336	0.770	0.300	1.025	5.647	0.307	1.693
14	0.802	0.235	0.817	0.981	1.019	0.406	1.594	0.398	1.563	3.407	0.763	0.294	1.118	5.696	0.328	1.672
15	0.775	0.223	0.789	0.982	1.018	0.428	1.572	0.421	1.544	3.472	0.756	0.288	1.203	5.740	0.347	1.653
16	0.750	0.212	0.763	0.983	1.017	0.448	1.552	0.440	1.527	3.532	0.750	0.283	1.282	5.782	0.363	1.637
17	0.728	0.203	0.739	0.985	1.016	0.466	1.534	0.459	1.510	3.588	0.744	0.279	1.356	5.820	0.378	1.622
18	0.707	0.194	0.718	0.985	1.015	0.482	1.518	0.475	1.496	3.640	0.739	0.275	1.424	5.856	0.391	1.609
19	0.688	0.187	0.698	0.986	1.014	0.497	1.503	0.490	1.483	3.689	0.733	0.271	1.489	5.889	0.404	1.596
20	0.671	0.180	0.680	0.987	1.013	0.510	1.490	0.503	1.470	3.735	0.729	0.268	1.549	5.921	0.415	1.585
21	0.655	0.173	0.663	0.988	1.013	0.523	1.477	0.516	1.459	3.778	0.724	0.265	1.606	5.951	0.425	1.575
22	0.640	0.167	0.647	0.988	1.012	0.534	1.466	0.528	1.448	3.819	0.720	0.262	1.660	5.979	0.435	1.565
23	0.626	0.162	0.633	0.989	1.011	0.545	1.455	0.539	1.438	3.858	0.716	0.259	1.711	6.006	0.443	1.557
24	0.612	0.157	0.619	0.989	1.011	0.555	1.445	0.549	1.429	3.895	0.712	0.257	1.759	6.032	0.452	1.548
25	0.600	0.153	0.606	0.990	1.010	0.565	1.435	0.559	1.420	3.931	0.708	0.254	1.805	6.056	0.459	1.541

Anexo 3:

CL(n, alpha)

```
#####Función cálculo de CL para 2D

CL<-function(n,alpha){
  fin=function(t){
    g=function(t,y){return(n*t-(n-2)*log(n-1)-y+(n-2)*log(y))}
    h=function(t,y){return(sqrt(g(t,y)*(g(t,y)>=0)))}
    f=function(y){return(dchisq(y,df=n-1)*(2*pnorm(h(t,y))-1))}
    return(1-integrate(f,lower=0,upper=100)$value-alpha)
  }
  return(nleqslv(x=3,fin)$x)
}
```

Anexo 4:

musigma(data,n,k,alpha,newdata)

```
#####Función gráfico de control Musigma
musigma<-function(datas,n,k,alpha=0.01,newdata){
#Datos muestrales
MeanData<-mean(datas[1:k,])
SdData<-sd(datas[1:k,])
Xj<-(datas[1:k,]-MeanData)/(SdData)
i<-1
Ematrix<-matrix(nrow=k)
for (i in 1:k){
  Ematrix[i,]<-mean(Xj[i,])
}

i<-1
Tmatrix<-matrix(nrow=k)
for (i in 1:k){
  Tmatrix[i]=sd(Xj[i,])
}

i<-1
Nmatrix<-matrix(nrow=k)
for(i in 1:k){

  Nmatrix[i]=((1-1/n)*(Tmatrix[i])^2-2*(1-2/5)*log(Tmatrix[i]))
}
i<-1
Squarematrix<-matrix(nrow=k)
for(i in 1:k){
Squarematrix[i]=(Nmatrix[i])+(Ematrix[i])^2
}
MuestralMatrix<-cbind(datas[1:k,],Ematrix,Tmatrix,Nmatrix,Squarematrix)
colnames(MuestralMatrix)<-c("sample 1","sample 2", "sample 3", "sample 4",
"sample 5","E","t","N","N^2+E^2")
#####Límites de control
#Obtención de C(alpha,n)
limit<-CL(n,alpha)
#####Gráfico de los valores (E,N)muestrales
par(bg="grey")
plot(1:k,Squarematrix,ann=FALSE,type="b", ylim=c(0,limit+1))
title(main="Gráfico de control de proceso 2D", xlab="Número de
muestra",ylab="Estadístico:T=?^2+?^2")
rect(par("usr")[1], par("usr")[3], par("usr")[2], par("usr")[4], col="white")
points(1:k,Squarematrix,ann=FALSE,type="b")
abline(h=limit, v=k,lty=2)
mtext("Datos históricos", side=3,adj=0)
```

```

text(1,limit+0.15,"LC",)
abline(h=limit, col="red",lty=2)
i<-1
for (i in 1:k){
  if(Squarematrix[i]>limit){
    points(i,Squarematrix[i], col="red")
  }
}

####New data
NewData<-rbind(datas,newdata)
i<-1
NewXj<-(newdata-MeanData)/(SdData)

NewEmatrix<-matrix(nrow=nrow(newdata))
for (i in 1:nrow(newdata)){
  NewEmatrix[i]<-mean(NewXj[i,])
}

i<-1
NewTmatrix<-matrix(nrow=nrow(newdata))
for (i in 1:nrow(newdata)){
  NewTmatrix[i,]=sd(NewXj[i,])
}

i<-1
NewNmatrix<-matrix(nrow=nrow(newdata))
for(i in 1:nrow(newdata)){
  NewNmatrix[i]=((1-1/n)*(NewTmatrix[i])^2-2*(1-2/n)*log(NewTmatrix[i]))
}
i<-1
NewSquareMatrix<-matrix(nrow=nrow(newdata))
for(i in 1:nrow(newdata)){
  NewSquareMatrix[i]=(NewNmatrix[i])+(NewEmatrix[i])^2
}
NewXj<-cbind(newdata,NewEmatrix,NewNmatrix,NewTmatrix,NewSquareMatrix)
MuestralMatrix<-rbind(MuestralMatrix,NewXj)
par(bg="grey")
plot(1:nrow(MuestralMatrix),MuestralMatrix[,9],ann=FALSE, type="b")
title(main="Gráfico de control de proceso 2D ", xlab="Número de
muestra",ylab="Estadístico: T=?^2+?^2")
rect(par("usr")[1], par("usr")[3], par("usr")[2], par("usr")[4], col="white")
points(1:nrow(MuestralMatrix),MuestralMatrix[,9],ann=FALSE, type="b")

```



```
abline(h=limit, v=k,lty=2)
mtext("Datos históricos", side=3,adj=0)
mtext("Nuevos Datos", side=3, adj=1)
text(1,limit+0.15,"LC",)
i<-1
for (i in 1:nrow(MuestralMatrix)){
  if(MuestralMatrix[i,9]>limit){
    points(i,MuestralMatrix[i,9], col="red")
    print(sprintf("La muestra %i está fuera de control",i))
  }
}
}
```

Anexo 5:

CLrango(m,n,alpha)

```
#####Función LC para SC

CLrango<-function(m,n,alpha){

  if(m==30){
    C<-
    matrix(c(5,11,25,3.99,4.09,3.78,4.26,4.30,4.01,4.48,4.45,4.18),ncol
    =4)
  }else if(m==50){ C<-
    matrix(c(5,11,25,4.62,4.38,4.25,4.97,4.63,4.50,5.25,4.80,4.70),ncol
    =4)

  }else if (m==100){ C<-
    matrix(c(5,11,25,5.22,4.79,4.73,5.64,5.11,5.03,5.98,5.34,5.25),ncol
    =4)
  }else if (m==125){C<-
    matrix(c(5,11,25,5.33,4.90,4.84,5.77,5.25,5.14,6.10,5.50,5.37),ncol
    =4)
  }else if (m==150){ C<-
    matrix(c(5,11,25,5.44,5.01,4.94,5.90,5.38,5.25,6.25,5.67,5.49),
    ncol=4)
  }
  colnames(C)<-c("n", " a=0.004", " a=0.0027", " a=0.002")
  if(alpha==0.004){ h<-2}
  else if(alpha==0.0027){ h<-3}
  else if(alpha==0.002){ h<-4}
  else {printf("alpha no válido")}
  if(n==5){v<-1}
  else if(n==11){ v<-2}
  else if(n==25){ v<-3}
  print(C[v,h])
}
```

Anexo 6:

rango(sample,m,n,alpha,test)

```
rango<-function(sample,m,n,alpha,test){
  ##Matriz datos muestrales
  Xm<-matrix(sample,nrow=m)
  p<-((2*((n+m)^2-4))/((2*(n+m)+1)*(8*(n+m)+11))-1
  cucconimuestral<-matrix(nrow=nrow(sample))
  Xm<-matrix(sample,nrow=m)
  j<-1
  for(j in 1:nrow(sample)){
    Ujn<-matrix(sample[j,],nrow=n)
    ranks<-cbind(sort(rbind(Xm,Ujn)),rank(sort(rbind(Xm,Ujn))))
    T1<-0
    S1<-0
    i<-1
    for(i in 1:n){
      h<-1
      for(h in 1:(m+n)){
        if(ranks[h,1]==Ujn[i]){rankjn<-ranks[h,2]}
      }
    }
    T1<-T1+rankjn
    S1<-S1+rankjn*rankjn

    S2<-n*((n+m)+1)^2-2*((n+m)+1)*T1+S1
    U<-(6*S1-
n*(n+m+1)*(2*(n+m)+1))/sqrt((m*n)/5*(n+m+1)*(2*(n+m)+1)*(8*(n+m)+11))
    V<-(6*S2-
n*(n+m+1)*(2*(n+m)+1))/sqrt((m*n)/5*(n+m+1)*(2*(n+m)+1)*(8*(n+m)+11))
    C<-(U^2+V^2-2*p*U*V)/(2*(1-p^2))
    cucconimuestral[j]<-C
  }

  #####Límite de control

  H<-CLrango(m,n,alpha)
  plot((1:nrow(sample)),cucconimuestral, ann=FALSE, type="b", ylim=c(0,
H+0.5))
  ##Gráfico del estadístico de Cucconi
  title(main= "Gráfico de control de rangos de Cucconi", xlab="Número de
muestra",ylab="Estadístico C")
  rect(par("usr")[1], par("usr")[3], par("usr")[2], par("usr")[4], col="white")
  points((1:nrow(sample)),cucconimuestral, ann=FALSE, type="b")
  abline(h=H, col="red", lty=2)
  mtext("Datos históricos", side=3,adj=0)
```

```

text(1,H+0.15, "LC")
i<-1
out<-0
for (i in 1:nrow(cucconimuestral)){
  if(cucconimuestral[i]>6.1){
    points(i,cucconimuestral[i], col="red")
    out<-1
    print(out)
  }
}

##Matriz de nuevas observaciones
cucconi<-matrix(nrow=nrow(test))
j<-1
for(j in 1:nrow(test)){
  Yjn<-matrix(test[j,],nrow=n)
  ranks<-cbind(sort(rbind(Xm,Yjn)),rank(sort(rbind(Xm,Yjn))))
  T1<-0
  S1<-0
  for(i in 1:n){
    h<-1
    for(h in 1:(m+n)){
      if(ranks[h,1]==Yjn[i]){rankjn<-ranks[h,2]}
    }
    T1<-T1+rankjn
    S1<-S1+rankjn*rankjn
  }
  S2<-n*((n+m)+1)^2-2*((n+m)+1)*T1+S1

  U<-(6*S1-
n*(n+m+1)*(2*(n+m)+1))/sqrt((m*n)/5*(n+m+1)*(2*(n+m)+1)*(8*(n+m)+11))

  V<-(6*S2-
n*(n+m+1)*(2*(n+m)+1))/sqrt((m*n)/5*(n+m+1)*(2*(n+m)+1)*(8*(n+m)+11))

  C<-(U^2+V^2-2*p*U*V)/(2*(1-p^2))

  cucconi[j]<-C
}
cucconi<-rbind(cucconimuestral,cucconi)
par(bg="grey")
plot((1:nrow(cucconi)),cucconi, ann=FALSE, type="b")
##Gráfico del estadístico de Cucconi

```

```
title(main= "Gráfico de control Shewhart-Cucconi", xlab="Número de  
muestra",ylab="Estadístico C")  
rect(par("usr")[1], par("usr")[3], par("usr")[2], par("usr")[4], col="white")  
points((1:nrow(cucconi)),cucconi, ann=FALSE, type="b")  
abline(h=H,v=m/n, lty=2)  
mtext("Datos históricos", side=3,adj=0)  
mtext("Nuevos Datos", side=3, adj=1)  
text(1,H+0.2, "LC")  
i<-1  
for (i in 1:nrow(cucconi)){  
  if(cucconi[i]>6.1){  
    points(i,cucconi[i], col="red")  
    print(sprintf("La muestra %i está fuera de control",i))  
  }  
}
```

Anexo 7:

Función para el cálculo del ARL del gráfico 2-D (en este caso con distribución normal)

ARLmusigma(n,k,alpha,mean1,sd1)

```
ARLmusigma<-function(n,k,alpha=0.01,mean1,sd1){
#Datos muestrales
q<-1
AR0<-0
suma<-0
count<-0
for(q in 1:1000){
continuess<-0
while(continuess==0){
datas<-matrix(rnorm(100,0,1),ncol=n)

MeanData<-mean(datas[1:k,])
SdData<-sd(datas[1:k,])
Xj<-(datas[1:k,]-MeanData)/(SdData)
i<-1
Ematrix<-matrix(nrow=k)
for (i in 1:k){
Ematrix[i,]<-mean(Xj[i,])
}

i<-1
Tmatrix<-matrix(nrow=k)
for (i in 1:k){
Tmatrix[i]=sd(Xj[i,])
}
i<-1
Nmatrix<-matrix(nrow=k)
for(i in 1:k){

Nmatrix[i]=((1-1/n)*(Tmatrix[i])^2-2*(1-2/5)*log(Tmatrix[i]))
}
i<-1
Squarematrix<-matrix(nrow=k)
for(i in 1:k){
Squarematrix[i]=(Nmatrix[i])+(Ematrix[i])^2
}
MuestralMatrix<-cbind(datas[1:k,],Ematrix,Tmatrix,Nmatrix,Squarematrix)
colnames(MuestralMatrix)<-c("sample 1","sample 2", "sample 3", "sample 4",
"sample 5","E","t","N","N^2+E^2")

#####Gráfico de los valores (E,N)muestrales
```

```

plot(1:k,Squarematrix,ann=FALSE,type="b")

title(main= "Gráfico de control de proceso usando media y varianza", xlab="Número
de muestra",ylab="Estadístico:E^2+N^2")

#####Límites de control
#Obtención de C(alpha,n)

limit<-CL(n,alpha)
#####Construcción de los límites de control

abline(h=limit, col="red",lty=2)
i<-1
continuess<-1
for (i in 1:k){
  if(Squarematrix[i]>limit){
    points(i,Squarematrix[i], col="red")
    continuess<-0
  }
}

}

####New data
i<-1
Squarematrix<-matrix(nrow=k)
for(i in 1:k){
  Squarematrix[i]=(Nmatrix[i])+(Ematrix[i])^2
}
MuestralMatrix<-cbind(datas[1:k,],Ematrix,Tmatrix,Nmatrix,Squarematrix)
colnames(MuestralMatrix)<-c("sample 1","sample 2", "sample 3", "sample 4",
"sample 5","E","t","N","N^2+E^2")
continue<-1
suma<-0
while(continue==1){
  var2<-matrix(rnorm(n,mean1,sd1),ncol=n)
  NewData<-rbind(datas,var2)
  i<-1
  NewXj<-(var2-MeanData)/(SdData)

  NewEmatrix<-matrix(nrow=nrow(var2))

```

```

for (i in 1:nrow(var2)){
  NewEmatrix[i]<-mean(NewXj[i,])
}
i<-1
NewTmatrix<-matrix(nrow=nrow(var2))
for (i in 1:nrow(var2)){
  NewTmatrix[i,]=sd(NewXj[i,])
}
i<-1
NewNmatrix<-matrix(nrow=nrow(var2))
for(i in 1:nrow(var2)){
  NewNmatrix[i]=((1-1/n)*(NewTmatrix[i])^2-2*(1-2/n)*log(NewTmatrix[i]))
}
i<-1
NewSquareMatrix<-matrix(nrow=nrow(var2))
for(i in 1:nrow(var2)){
  NewSquareMatrix[i]=(NewNmatrix[i])+(NewEmatrix[i])^2
}
NewXj<-cbind(var2,NewEmatrix,NewNmatrix,NewTmatrix,NewSquareMatrix)
MuestralMatrix<-rbind(MuestralMatrix,NewXj)
plot(1:nrow(MuestralMatrix),MuestralMatrix[,9],ann=FALSE, type="b")
title(main= "Gráfico de control 2D", xlab="Número de muestra",ylab="Estadístico:
T=?^2+?^2")
abline(h=limit, col="red",lty=2)
abline(v=k, lty=2)
i<-1
for (i in 1:nrow(MuestralMatrix)){
  if(MuestralMatrix[i,9]>limit){
    points(i,MuestralMatrix[i,9], col="red")

    continue<-0
  }

}
suma<-suma+1
}
AR0<-AR0+suma
count<-count+1
print(count)
}
print("AR0")
print(AR0/1000)
}

```


Anexo 8:

Función cálculo del ARL para el gráfico SC (en este caso para distribución de Laplace)

```
ARLrango<-function(m,n,alpha,mean1,sd1){
  z<-1
  AR0<-0
  count<-0
  for(z in 1:1000){
    out<-1
    while(out==1){
      p<-(2*((n+m)^2-4))/((2*(n+m)+1)*(8*(n+m)+11))-1

      sample<-matrix(rdoubles(100,0,1),ncol=n)
      cucconimuestral<-matrix(nrow=nrow(sample))
      Xm<-matrix(sample,nrow=m)

      j<-1
      for(j in 1:nrow(sample)){
        Ujn<-matrix(sample[j,],nrow=n)
        ranks<-cbind(sort(rbind(Xm,Ujn)),rank(sort(rbind(Xm,Ujn))))
        T1<-0
        S1<-0
        i<-1
        for(i in 1:n){
          h<-1
          for(h in 1:(m+n)){
            if(ranks[h,1]==Ujn[i]){rankjn<-ranks[h,2]}
          }

          T1<-T1+rankjn
          S1<-S1+rankjn*rankjn

          S2<-n*((n+m)+1)^2-2*((n+m)+1)*T1+S1

          U<-(6*S1-
n*(n+m+1)*(2*(n+m)+1))/sqrt((m*n)/5*(n+m+1)*(2*(n+m)+1)*(8*(n+m)+11))

          V<-(6*S2-
n*(n+m+1)*(2*(n+m)+1))/sqrt((m*n)/5*(n+m+1)*(2*(n+m)+1)*(8*(n+m)+11))

          C<-(U^2+V^2-2*p*U*V)/(2*(1-p^2))

          cucconimuestral[j]<-C
        }
      }
    }
  }
}
```

```

    }
}
H<-CLrango(m,n,alpha)
plot((1:nrow(sample)),cucconimuestral, ann=FALSE, type="b")
  title(main= "Gráfico de control de rangos de Cucconi", xlab="Número de
muestra",ylab="Estadístico C")
  abline(h=H, col="red", lty=2)
  i<-1
  out<-0
  for (i in 1:nrow(cucconimuestral)){
    if(cucconimuestral[i]>H){
      points(i,cucconimuestral[i], col="red")
      out<-1
      print(out)
    }
  }
  contador<-0
  seguir<-1
  while(seguir==1){

    test<-matrix(rdoubles(n,mean1,sd1),ncol=n)
    Yjn<-matrix(test,nrow=n)
    ranks<-cbind(sort(rbind(Xm,Yjn)),rank(sort(rbind(Xm,Yjn))))
    T1<-0
    S1<-0
    i<-1
    for(i in 1:n){
      h<-1
      for(h in 1:(m+n)){
        if(ranks[h,1]==Yjn[i]){rankjn<-ranks[h,2]
        }
      }

      T1<-T1+rankjn
      S1<-S1+rankjn*rankjn

      S2<-n*((n+m)+1)^2-2*((n+m)+1)*T1+S1

      U<-(6*S1-
n*(n+m+1)*(2*(n+m)+1))/sqrt((m*n)/5*(n+m+1)*(2*(n+m)+1)*(8*(n+m)+11))

      V<-(6*S2-
n*(n+m+1)*(2*(n+m)+1))/sqrt((m*n)/5*(n+m+1)*(2*(n+m)+1)*(8*(n+m)+11))

      C<-(U^2+V^2-2*p*U*V)/(2*(1-p^2))

      cucconi2<-C

```

```
    }  
    cucconi<-rbind(cucconimuestral,cucconi2)  
    plot(1:(nrow(cucconi)),cucconi, ann=FALSE, type="b")  
    title(main= "Gráfico de control de rangos de Cucconi", xlab="Número de  
muestra",ylab="Estadístico C")  
  
    abline(h=H ,col="red", lty=2)  
    i<-1  
    for (i in nrow(cucconimuestral):nrow(cucconi)){  
        if(cucconi[i]>H){  
            points(i,cucconi[i], col="red")  
            seguir<-0  
        }  
    }  
    contador<-contador+1  
    }  
    count<-count+1  
    }  
    AR0<-AR0+contador  
    print(count)  
    }  
    print(AR0/1000)  
    }
```



TFG: Gráficos de control univariantes según el promedio y la desviación típica del rango